Res. Lett. Inf. Math. Sci., 2006, Vol. **10**, pp 17-48
Available online at http://iims.massey.ac.nz/research/letters/

Variances of first passage times in a Markov chain with applications to mixing times

J. J. HUNTER

Institute of Information & Mathematical Sciences
Massey University at Albany, Auckland, New Zealand

In an earlier paper the author introduced the statistic $\eta_i = \sum_{j=1}^{m} m_{ij}\pi_j$ as a measure of the "mixing time" or "time to stationarity" in a finite irreducible discrete time Markov chain with stationary distribution $\{p_j\}$ and m_{ij} as the mean first passage time from state i to state j of the Markov chain. This was shown to be independent of the initial state i with $\eta_i = \eta$ for all i, minimal in the case of a periodic chain, yet can be arbitrarily large in a variety of situations. In this paper we explore the variance of the mixing time v_i, starting in state i. The v_i, are shown to depend on i and an exploration of recommended starting states, given knowledge of the transition probabilities, is considered. As a preamble, a study of the computation of second moments of the mixing times, $m_{ij}^{(2)}$, and the variance of the first passage times, in a discrete time Markov chain is carried out leading to some new results.

1 Introduction

Let $P = [p_{ij}]$ be the transition matrix of a finite irreducible, discrete time Markov chain $\{X_n\}$, $(n \geq 0)$, with state space $S = \{1, 2,\ldots, m\}$. It is well known that such Markov chains have a unique stationary distribution $\{\pi_j\}$, $(1 \leq j \leq m)$, that, in the case of a regular (finite, irreducible and aperiodic) chain, is also the limiting distribution of the Markov chain ([10, Theorem 7.1.2]). Let $\boldsymbol{\pi}^T = (\pi_1, \pi_2, \ldots, \pi_m)$ be the stationary probability vector of the Markov chain.

Let T_{ij} be the first passage time random variable from state i to state j, i.e. $T_{ij} = min\{n \geq 1$ *such that* $X_n = j$ *given that* $X_0 = i\}$, so that T_{ii} is the "first return to state i". Let $M = [m_{ij}]$ be the matrix of the mean first passage times from state i to state j, i.e. $m_{ij} = E[T_{ij} \mid X_0 = i]$ for all $i, j \in S$.

The irreducibility of the Markov chain ensures that the T_{ij} are all proper random variables ([9, Theorem 5.3.6]), and under the finite state space restriction, all the moments of T_{ij} are finite, ([10, Theorem 7.3.1]). (It is possible, in the presence of null states in the case of an infinite state space for $m_{ii} = +\infty$.)

Definition 1.1. *(T, the "time to mixing" in a Markov chain)*
Let Y be a random variable whose probability distribution is the stationary distribution $\{\pi_j\}$. *We shall say that the Markov chain* $\{X_n\}$, *"reaches stationarity", or achieves "mixing", at time* $T = k$, *when* $X_k = Y$ *for the smallest such* $k \geq 1$.

Thus, we first sample from the stationary distribution $\{\pi_j\}$ to determine a value of the random variable Y, say $Y = j$. We then observe the Markov chain, starting at a given state i and achieve "mixing" at time $T = n$ when $X_n = j$ for the first such $n \geq 1$. i.e., conditional upon $Y = j$, $T = T_{ij}$, the first passage time from state i to state j, (or return to state j when $i = j$). The finite state space restriction, under irreducibility conditions, ensures the finiteness of the "mixing time" (a.s), with a finite expectation.

Let us define $\eta_i^{(k)} = E[T^k \mid X_0 = i]$ as the *k-th* moment of the mixing time, starting in state i. Also let $m_{ij}^{(k)}$ be the *k-th* moment of the first passage time from state i to state j in the Markov chain.

Theorem 1.1. *For all i,* $\eta_i^{(k)} = \sum_{j=1}^{m} m_{ij}^{(k)}\pi_j$.

Proof. $E[T^k] = E_Y(E[T^k \mid Y]) = \sum_{j=1}^{m} E[T^k \mid Y = j]\, P[Y = j]$

$$= \sum_{j=1}^{m} E[T_{ij}^k \mid X_0 = i]\pi_j = \sum_{j=1}^{m} m_{ij}^{(k)}\pi_j. \quad \square$$

In [3] we explored the properties of $\eta_i = \sum_{j=1}^{m} m_{ij}\pi_j$, the expected time to mixing starting at state i, and showed that $\eta_i = \eta$, a constant independent of i, the starting state. Since the average time to mixing does not depend on which state the Markov chain commences, we pose the following question: Is there some particular state that could be more desirable as a starting point? The distributions of the mixing times, starting from different states could possibly have some widely different characteristics, even though they have the same mean. In this paper we derive expressions for the variances v_i, of the mixing times starting in each state i and then explore the possibility of choosing some particular state that say minimises the variances v_i of the mixing times over the state space S. The existence of such a state would certainly be desirable from the view of eliminating any widely varying mixing times providing some tighter bounds on the time to mixing.

Observe that $v_i = \text{var}\,[T \mid X_0 = i] = \eta_i^{(2)} - \eta^2 = \sum_{j=1}^{m} m_{ij}^{(2)}\pi_j - \eta^2$, thus we first need to explore the derivation of the first two moments of the first passage times.

In Section 2 we develop some new results for the second moment and the variances of the first passage times in a Markov chain based upon the use of generalized matrix inverses in solving systems of linear equations. These results are used in Section 3 to derive expressions for the variances of the mixing times starting in a particular state. In Section 4 we apply the results to general two-state and three-state Markov chains, in the latter model focusing on a variety of special cases. The paper contains many new results and opens up an area of research, in particular the application of the results to more general Markov chains.

2 Moments of the first passage times

Let $M = [m_{ij}]$ and $M^{(2)} = [m_{ij}^{(2)}]$ be the first and second moments of the first passage times from state *i* to state *j* in an irreducible finite Markov chain with transition matrix P.

In earlier papers ([6], [11], [12]), we have shown that generalized inverses (g-inverses) of the Markovian kernel $I - P$ have useful properties in determining the mean first passage times.

Theorem 2.1. *M satisfies the matrix equation*

$$(I - P)M = E - PM_d, \tag{2.1}$$

where, if $\boldsymbol{e}^T = (1, 1, ..., 1)$, $E = \boldsymbol{ee}^T = [1]$, $M_d = [\delta_{ij}m_{ij}]$, *a diagonal matrix with elements the diagonal elements of M.*

Proof: Equation (2.1) is well known. (See for example, [10, Corollary 7.3.3B], [11, Section 5.1], [13, Theorem 4.4.4]). ❑

Lemma 2.2. *If X is an arbitrary square matrix, and Λ is a diagonal matrix,*

$$(XE)_d = (X\Pi)_d D, \quad (X\Lambda)_d = X_d\Lambda, \quad E\Pi_d = \Pi\,. \tag{2.2}$$

Proof: These results are well known. See [10, Lemma 7.3.5]. ❑

Corollary 2.1.1. $M_d = (\Pi_d)^{-1} \equiv D$, where $\Pi = \boldsymbol{e\pi}^T$.

Proof: This result is well known since $m_{ii} = 1/\pi_i$. It also follows by multiplying both sides of (2.1) by Π. Note that $\Pi P = \Pi$ so that $\Pi E = E = \Pi M_d$ and the result follows by taking diagonal elements and noting that $E_d = I = \Pi_d M_d$. ❑

It is well know that the solution of equations of the form of Eqn. (2.1) can be effected using g-inverses of I – P, (see e.g. [10] and [11]). Any g-inverse of $I - P$ has the form

$$G = [I - P + \boldsymbol{tu}^T]^{-1} + \boldsymbol{ef}^T + \boldsymbol{g\pi}^T$$

where $\boldsymbol{u}^T\boldsymbol{e} \neq 0$, $\boldsymbol{\pi}^T\boldsymbol{t} \neq 0$ and $\boldsymbol{f}$ and $\boldsymbol{g}$ are arbitrary vectors.

Theorem 2.3. *If G is any g-inverse of I – P, then*

$$M = [G\Pi - E(G\Pi)_d + I - G + EG_d]D. \tag{2.3}$$

Proof: Result (2.3) appears in [10, Theorem 7.3.6] and [11, Theorem 5.1]. ❑

By choosing special g-inverses, Eqn.(2.3) can be simplified. First note the following consequence of Theorem 2.3.

Corollary 2.3.1.

$$GE - E(G\Pi)_dD = M - [I - G + EG_d]D. \tag{2.4}$$

Proof: Result (2.4) follows from Eqn.(2.3) by noting that $\Pi D = E$. ❑

If $A = [a_{ij}]$ is a matrix, define $a_{i\bullet} = \sum_{j=1}^{m} a_{ij}$.

Corollary 2.3.2. *Under any of the following three equivalent conditions,*

(i) $G\boldsymbol{e} = g\boldsymbol{e}$, *g a constant,*
(ii) $GE - E(G\Pi)_dD = 0$,
(iii) $G\Pi - E(G\Pi)_d = 0$,

$$M = [I - G + EG_d]D. \tag{2.5}$$

Proof: Condition *(i)* $G\boldsymbol{e} = g\boldsymbol{e}$ implies $GE = G\boldsymbol{e}\boldsymbol{e}^T = g\boldsymbol{e}\boldsymbol{e}^T = gE$ and $E(G\Pi)_dD = E(G\boldsymbol{e}\boldsymbol{\pi}^T)_dD = E(g\boldsymbol{e}\boldsymbol{\pi}^T)_dD = gE\Pi_dD = gE$ leading to condition *(ii)*. Since $\Pi D = E$ substitution in *(ii)* and post-multiplication by D^{-1} leads to condition *(iii)*. Further under condition *(iii)*, $G\Pi = G\boldsymbol{e}\boldsymbol{\pi}^T = \boldsymbol{e}\boldsymbol{e}^T(G\Pi)_d$. Post-multiplication by $\boldsymbol{e}$, since $\boldsymbol{\pi}^T\boldsymbol{e} = 1$, yields $G\boldsymbol{e} = \boldsymbol{e}\boldsymbol{e}^T(G\Pi)_d\boldsymbol{e} = g\boldsymbol{e}$ where $g = \boldsymbol{e}^T(G\Pi)_d\boldsymbol{e}$, a constant $(= \sum_{j=1}^{m} g_{j\bullet}\pi_j)$. Thus *(i)* $\Rightarrow$ *(ii)* $\Rightarrow$ *(iii)* $\Rightarrow$ *(i)* and the conditions are equivalent. Result (2.5) follows from Eqn.(2.4) under condition *(ii)*. Note that condition *(i)* implies that $g_{j\bullet} = g$ for all j. ❑

From [7, Cor 3.1.1] the conditions of Corollary 2.3.2 also imply that if $G = H + \boldsymbol{e}\boldsymbol{f}^T + \boldsymbol{g}\boldsymbol{\pi}^T$ then $M = [I - H + EH_d]D$ if and only if $H = [I - P + \boldsymbol{e}\boldsymbol{u}^T]^{-1}$ for some $\boldsymbol{u}$ such that $\boldsymbol{u}^T\mathbf{e} \neq 0$.

If we take $G = H + \boldsymbol{e}\boldsymbol{f}^T$ then $G\boldsymbol{e} = g\boldsymbol{e}$ with $g = 1 + \boldsymbol{f}^T\boldsymbol{e}$. Special cases for Eqn.(2.5) are $G = Z$, Kemeny and Snell's fundamental matrix $Z = [I - P - \Pi]^{-1}$ (since $Z\boldsymbol{e} = \boldsymbol{e}$ and g = 1) as given initially in [13, Theorem 4.4.7] and $G = T = Z - \Pi$, Meyer's group inverse of $I - P$, (with $T\boldsymbol{e} = \boldsymbol{0}$ and g = 0) as given in [14, Theorem 3.3].

Elemental expressions for the m_{ij} follow from Theorem 2.3 as follows.

Corollary 2.3.3. *If* $G = [\, g_{ij}]$ *then*

$$m_{ij} = ([g_{jj} - g_{ij} + \delta_{ij}]/\pi_j) + (g_{i\cdot} - g_{j\cdot}), \text{ for all } i, j. \tag{2.5}$$

Further, when $G\boldsymbol{e} = g\boldsymbol{e}$,

$$m_{ij} = [g_{jj} - g_{ij} + \delta_{ij}]/\pi_j \quad \text{for all } i, j. \tag{2.6}$$

We parallel the development, given above for M, for $M^{(2)}$.

Theorem 2.4. $M^{(2)}$ *satisfies the matrix equation*

$$(I - P)M^{(2)} = E + 2P(M - M_d) - PM_d^{(2)}. \tag{2.7}$$

Proof: This result is well known. See [10, Theorem 7.3.10], [13, proof of Theorem 4.5.1]. ❑

Corollary 2.4.1.

$$M_d^{(2)} = 2D(\Pi M)_d - D. \tag{2.8}$$

Proof: Pre-multiplication of both sides of Eqn. (2.7) with Π, noting that $M_d = D$, $\Pi = \Pi P$, and $\Pi E = E$, yields

$$\Pi M_d^{(2)} = E + 2\Pi(M - D).$$

Now take diagonal elements, using Lemma 2.2, to obtain

$$\Pi_d M_d^{(2)} = I + 2(\Pi M)_d - 2\Pi_d D.$$

Eqn.(2.8) follows, from Corollary 2.1, since $\Pi_d^{-1} = D$. ❑

Corollary 2.4.2. *If* G *is any g-inverse of* $I - P$,

$$M_d^{(2)} = D + 2D\{(I - \Pi)G(I - \Pi)\}_d D. \tag{2.9}$$

Proof: From Eqn.(2.3), noting that $\Pi E = E$, it is easily seen that

$$\Pi M - M = [(\mathrm{I} - \Pi)G(I - \Pi) + \Pi - I]D.$$

Taking the diagonal elements, and noting that $M_d = D$ and $\Pi_d D = I$, yields

$$(\Pi M)_d = \{(I - \Pi)G(I - \Pi)\}_d D + I. \tag{2.10}$$

Eqn (2.9) now follows from Eqn.(2.7). ❑

Corollary 2.4.3. *If* $G\boldsymbol{e} = g\boldsymbol{e}$,

$$M_d^{(2)} = D + 2D\{(I - \Pi)G\}_d D. \tag{2.11}$$

In particular,

$$M_d^{(2)} = D + 2DT_dD, \quad (2.12)$$
$$= 2DZ_dD - D. \quad (2.13)$$

Proof: Eqn.(2.11) follows from Eqn.(2.9) by observing that $G\Pi = G\boldsymbol{e}\boldsymbol{\pi}^T = \boldsymbol{e}\boldsymbol{\pi}^T = \Pi$ and that $(I - \Pi)\Pi = 0$ since $\Pi^2 = \boldsymbol{e}\boldsymbol{\pi}^T\boldsymbol{e}\boldsymbol{\pi}^T = \boldsymbol{e}\boldsymbol{\pi}^T = \Pi$.

Eqn.(2.12) follows directly from Eqn.(2.9) as it has been shown ([11,Theorem 6.3]) that for all g-inverses G of I - P, $(I - \Pi)G(I - \Pi)$ is invariant and is in fact T, the group inverse of $I - P$.

Eqn.(2.13) follows directly from Eqn.(2.12) since $T = Z - \Pi$. ❑

Expression (2.9) is also given in [7, Eqn.(3.8)]; expression (2.12) is given in [14, proof Theorem 3.4] and [10, Eqn.7.3.25]; and expression (2.13) is given in [13, Theorem 4.5.2].

Closed form expresssions for $M^{(2)}$ have been derived by Kemeny and Snell [13] (using Z, the fundamental matrix), Meyer [14] (using T the group inverse) and Hunter [10], [11] and [12]. The only presentations using arbitrary g-inverses of $I - P$ are those of Hunter [10] and [11]. In [7] general techniques, using generalized inverses of $I - P$, for finding the higher moments of the first passage times were also discussed.

The results given by the following theorem are however new and offer possible computational advantages.

Theorem 2.5. *If G is any g-inverse of I – P, then*

$$M^{(2)} = 2[GM - E(GM)_d] + [I - G + EG_d][M_d^{(2)} + D] - M, \quad (2.14)$$
$$= 2[GM - E(GM)_d] + 2[I - G + EG_d]D(\Pi M)_d - M. \quad (2.15)$$

Proof: Eqn.(2.7) is of the form $AX = C$, where $A = I - P$, $X = M^{(2)}$ and C is known. Consistent equations of this form can be solved using any g-inverse of A, A^-, with the general solution given by $X = A^-B + (I - A^-A)U$, where U is an arbitrary matrix (see [11], Corollary 3.1.1]. (The consistency condition, $A^-AC = C$, can be shown to be satisfied.)

Thus the general solution of Eqn.(2.7), with G is any g-inverse of $I - P$, is given by

$$M^{(2)} = G[E + 2P(M - M_d) - PM_d^{(2)}] + [I - G(I - P)]U \quad (2.16)$$

where U is an arbitrary matrix. The arbitrariness of U can be eliminated by taking advantage of the knowledge of $M_d^{(2)}$. We first simplify Eqn.(2.16) by using Eqn. (2.7) to show that

$$GP(M - M_d) = GPM - GPD = GM - GE.$$

Secondly, for any g-inverse G of $I – P$, it can be shown that for some $\boldsymbol{\beta}^T$, (see [6, Eqn.(2.15)]) that

$$I - G(I - P) = \boldsymbol{e}\boldsymbol{\beta}^T. \tag{2.17}$$

Thus, defining $\boldsymbol{\beta}^T U = \boldsymbol{b}^T$ and noting that $\boldsymbol{e}\boldsymbol{b}^T$ can be expressed as EB, say, where B is a diagonal matrix, whose diagonal elements are those of the vector $\boldsymbol{b}^T$. Thus Eqn.(2.16) can be expressed as

$$M^{(2)} = 2GM - GE - GPM_d^{(2)} + EB. \tag{2.18}$$

We now determine B by taking the diagonal elements of Eqn.(2.18) and, using an appropriate expression for $M_d^{(2)}$, to obtain

$$B = M_d^{(2)} - 2(GM)_d + (G\Pi)_d D + (GP)_d M_d^{(2)}. \tag{2.19}$$

Substitution of B, from Eqn.(2.19) into Eqn.(2.18) yields

$$M^{(2)} = 2GM - 2E(GM)_d - GE + E(G\Pi)_d D + [E + E(GP)_d - GP]M_d^{(2)}. \tag{2.20}$$

Further simplification of Eqn.(2.20) is possible. From Eqn.(2.17) observe that

$$E(I - G + GP)_d = \boldsymbol{e}\boldsymbol{e}^T(\boldsymbol{e}\boldsymbol{\beta}^T)_d = \boldsymbol{e}\boldsymbol{\beta}^T = I - G + GP$$

yielding, after further refinement,

$$M^{(2)} = 2GM - 2E(GM)_d - GE + E(G\Pi)_d D + [I - G + EG_d]M_d^{(2)}. \tag{2.21}$$

We can now make use of Corollary 2.3.1 to obtain Eqn.(2.14).

Eqn.(2.15) follows from Corollary 2.4.1. ❑

Corollary 2.5.1 *If **G** is any g-inverse of I – P, such that G**e** = g**e**, then*

$$M^{(2)} = 2[GM - E(GM)_d] + MD^{-1}M_d^{(2)}. \tag{2.22}$$

Proof. When $G\boldsymbol{e} = g\boldsymbol{e}$, from Cor 2.3.2, $GE - E(G\Pi)_d D = 0$ and $[I - G + EG_d] = MD^{-1}$. Eqn.(2.22) now follows immediately from Eqn.(2.21). ❑

Corollary 2.5.2.

$$M^{(2)} = 2[ZM - E(ZM)_d] + M(2Z_d D - I), \tag{2.23}$$

$$= 2[TM - E(TM)_d] + M(2T_d D + I). \tag{2.24}$$

Proof. Eqns.(2.23) and (2.24) follow from Eqn.(2.22) making use of Eqns.(2.12) and (2.13). ❑

The original derivation of Eqn.(2.23) is due to Kemeny and Snell [13, Theorem 4.5.3] but was by indirect methods. A proof using Z along the lines given above (but not using arbitrary g-inverses) was given by the author [10,Theorem 7.3.10]. Meyer [14, Theorem 3.4], using T as a g-inverse of $I – P$, deduced expression (2.23).

Eqns.(2.23) and (2.24) have, in the past, provided the standard computational methods for finding $M^{(2)}$. Theorem 2.5 is new and has the added computational advantage that any g-inverse of $I - P$ can be used. This is particularly important since the previously used g-inverses (Z and T) both required and involved expressions for the stationary distribution of the Markov chain.

The efficient computation of the moments of first passage times in a Markov chain has attracted interest, see for example Heyman and Reeves [2], and Heyman and O'Leary [1]. Techniques for solving linear equations occurring in Markov chains involving Gaussian elimination, algorithmic methods, state reduction and hybrid methods have been popular but in many instances they have been based upon the closed form solutions given by Eqn.(2.5) for the mean first passage times, and Eqns.(2.23) and (2.24) for the second moments and variances, using either Z or T. In particular it was noted in [2] that there are potential accuracy problems in using such closed forms where it is pointed out that "the first thing that needs to be done is to compute Z. The majority of the work is to compute $\boldsymbol{\pi}$ and to do a matrix inversion." Further " There are three sources of numerical error. The first is the algorithm to compute $\boldsymbol{\pi}$. The second occurs in computing the inverse of $(I - P - \Pi)$; this matrix may have negative elements, and this can cause round-off errors when the inverse is evaluated. The third is the matrix multiplication in Eqn.(2.5); the matrix multiplying D may have negative elements. Now we consider the additional work to compute $M^{(2)}$, and the additional numerical errors that might occur. There are three matrix multiplications that are required, two of which involve at least one diagonal matrix. … In each of these multiplications there is a matrix with (possibly) negative elements, which may introduce round-off errors." In [1] it is noted that "Deriving means and variances of first passage times from either the fundamental matrix Z or the group generalized inverse T leads to a significant inaccuracy on the more difficult problems." The authors then conclude that "for this reason, it does not make sense to compute either the fundamental matrix or the group generalized inverse unless the individual elements of those matrices are of interest."

However the only closed form expressions utilized in the literature for finding M and $M^{(2)}$, thus far, have been expressions involving Z and T.

Corollary 2.5.1 provides a much simpler form for the computation of $M^{(2)}$ if one is prepared to restrict attention to the class of g-inverses of $I - P$ that have the property $G\boldsymbol{e} = g\boldsymbol{e}$. While Z and T both satisfy this restriction we can take advantage of much simpler forms of such g-inverses. We explore some further consequences of this observation in Example 2.1, below.

Elemental expressions for $m_{ij}^{(2)}$ can be found from Eqn.(2.14). If $A = [a_{ij}]$ then $EA_d = [a_{jj}]$.

Corollary 2.5.3. *If* $G = [g_{ij}]$ *then*

$$m_{ij}^{(2)} = 2\sum_{k=1}^{m}(g_{ik} - g_{jk})m_{kj} - m_{ij} + (\delta_{ij} - g_{ij} + g_{jj})(m_{jj}^{(2)} + m_{jj}). \tag{2.25}$$

Similarly, for inverses with the property that $G\boldsymbol{e} = g\boldsymbol{e}$, we have the following expression from Eqn(2.22).

Corollary 2.5.4. *If* $\boldsymbol{G}$ *is any g-inverse of* $I - P$, *such that* $G\boldsymbol{e} = g\boldsymbol{e}$, *then*

$$m_{ij}^{(2)} = 2\sum_{k=1}^{m}(g_{ik} - g_{jk})m_{kj} + m_{ij}m_{jj}^{(2)} \Big/ m_{jj}. \tag{2.26}$$

Eqns.(2.25) and (2.26) do not provide an explicit expressions for the $m_{jj}^{(2)}$. These can be derived from Eqn.(2.8).

Corollary 2.5.5.

$$m_{jj}^{(2)} + m_{jj} = 2m_{jj}\sum_{i=1}^{m}\pi_i m_{ij}. \tag{2.27}$$

To utilise Equation (2.27), simplified expressions of $\sum_{i=1}^{m}\pi_i m_{ij}$ are required.

Corollary 2.5.6. *If* $\boldsymbol{\alpha}^T \equiv (\alpha_1,...,\alpha_m)$ *where* $\alpha_j \equiv \sum_{i=1}^{m}\pi_i m_{ij}$, *then*

$$\boldsymbol{\alpha}^T = \boldsymbol{\pi}^T M = \boldsymbol{e}^T(\Pi M)_d. \tag{2.28}$$

Further, $\boldsymbol{\alpha} = (\Pi M)_d\boldsymbol{e}$ *and if G is any g-inverse of I – P, then*

$$\boldsymbol{\alpha} = (\boldsymbol{\pi}^T G\boldsymbol{e})\boldsymbol{e} - (GE)_d\boldsymbol{e} + \boldsymbol{e} - (\Pi G)_d D\boldsymbol{e} + G_d D\boldsymbol{e}. \tag{2.29}$$

In particular if $G\boldsymbol{e} = g\boldsymbol{e}$,

$$\boldsymbol{\alpha} = \boldsymbol{e} - (\Pi G)_d D\boldsymbol{e} + G_d D\boldsymbol{e}. \tag{2.30}$$

If Z is the fundamental matrix and T is the group inverse,

$$\boldsymbol{\alpha} = Z_d D\boldsymbol{e} = \boldsymbol{e} + T_d D\boldsymbol{e}. \tag{2.31}$$

Proof: The first expression of Eqn.(2.28) follows from the definition. Note that $\Pi M = \boldsymbol{e}\boldsymbol{\pi}^T M = \boldsymbol{e}\boldsymbol{\alpha}^T$, and consequently $\alpha_j = (\Pi M)_{jj} = (\Pi M)_{ij}$ for all *i, j.* This implies that $\boldsymbol{\alpha}^T = \boldsymbol{e}^T ?(\Pi M)_d$ and hence $\boldsymbol{\alpha} = (\Pi M)_d\boldsymbol{e}$.
From Eqn.(2.3), $\boldsymbol{\alpha}^T = \boldsymbol{\pi}^T M = \boldsymbol{\pi}^T[G\Pi - E(G\Pi)_d + I - G + EG_d]D.$ Simplification yields $\boldsymbol{\alpha}^T = (\boldsymbol{\pi}^T G\boldsymbol{e})\boldsymbol{e}^T - \boldsymbol{e}^T(GE)_d + \boldsymbol{e}^T - \boldsymbol{e}^T(\Pi G)_d D + \boldsymbol{e}^T G_d D,$ since $\Pi D = E = \boldsymbol{e}\boldsymbol{e}^T$, $\boldsymbol{\pi}^T E = \boldsymbol{e}^T$, $\boldsymbol{\pi}^T D = \boldsymbol{e}^T$, $\boldsymbol{\pi}^T G = \boldsymbol{e}^T(\Pi G)_d$, and from Eqn.(2.2), $(G\Pi)_d E = (GE)_d$.
Eqn.(2.29) follows by noting that if Λ is a diagonal matrix and $\boldsymbol{a}^T = \boldsymbol{e}^T\Lambda$ then $\boldsymbol{a} = \Lambda\boldsymbol{e}$.
An alternative derivation also follows from Eqn.(2.10) by noting that $\boldsymbol{\alpha} = (\Pi M)_d\boldsymbol{e}$ $= [(I - \Pi)G(I - \Pi)]_d D\boldsymbol{e} + \boldsymbol{e}.$ The equivalence follows by noting that $(\Pi G\Pi)_d D = (\boldsymbol{\pi}^T G\boldsymbol{e})I$.

When $G\boldsymbol{e} = g\boldsymbol{e}$, $(\boldsymbol{\pi}^T G\boldsymbol{e})\boldsymbol{e} - (G\Pi)_d D\boldsymbol{e} = g(\boldsymbol{\pi}^T\boldsymbol{e})\boldsymbol{e} - g\Pi_d D\boldsymbol{e} = g\boldsymbol{e} - g\boldsymbol{e} = \boldsymbol{0}$, and Eqn.(2.30) follows from Eqn.(2.29). Eqns.(2.31) follow from Eqn.(2.30) and the facts that $(\Pi Z)_d D = \Pi_d D = I$ and $T = Z - \Pi$. ❑

The transpose variant of Eqn.(2.31), using Z, has been derived earlier - see [10, Theorem 7.3.8], [13, Theorem 4.49]. The expressions using an arbitrary g-inverse G are new.

Elemental expressions for the α_j follow either from the expressions of Corollary 2.5.6 for $\boldsymbol{\alpha}^T$, or from the expressions of Corollary 2.3.3 for the m_{ij}.

Corollary 2.5.7. *If $G = [g_{ij}]$ is any g-inverse of $I - P$, then*

$$\alpha_j = \sum_{i=1}^m \pi_i m_{ij} = 1 + \sum_{i=1}^m \pi_i g_{i\bullet} - g_{j\bullet} + \left(g_{jj} - \sum_{i=1}^m \pi_i g_{ij}\right)\Big/\pi_j. \tag{2.32}$$

In particular if $G\boldsymbol{e} = g\boldsymbol{e}$ (i.e. $g_{i\bullet} = g$ for all i),

$$\alpha_j = 1 + \left(g_{jj} - \sum_{i=1}^m \pi_i g_{ij}\right)\Big/\pi_j\,. \tag{2.33}$$

Further if $Z = [z_{ij}]$ and $T = [t_{ij}]$,

$$\alpha_j = \left(z_{jj}/\pi_j\right) = 1 + \left(t_{jj}/\pi_j\right). \tag{2.34}$$

Some expressions in Cor 2.5.7 can be found directly by matrix-vector multiplications. For all $G = [g_{ij}]$, $\boldsymbol{\pi}^T G\boldsymbol{e} = \sum_{i=1}^m \pi_i g_{i\bullet} = \boldsymbol{e}^T(\Pi G)_d\boldsymbol{e} = tr(\Pi G) = tr(G\Pi)$, and $\boldsymbol{\pi}^T G\boldsymbol{e}_j = \sum_{i=1}^m \pi_i g_{ij}$.

Eqn.(2.34) is consistent with the observation from Eqn.(2.13) that

$$M_d^{(2)} + D = 2DZ_dD = [m_{jj}^{(2)} + m_{jj}] = 2\,[z_{jj}/(\pi_j)^2] = 2[\alpha_j m_{jj}].$$

The variances of the first passage times T_{ij} can be derived as

$$\text{var}[T_{ij}] = \text{var}\,[T_{ij} \mid X_0 = i] = m_{ij}^{(2)} - (m_{ij})^2.$$

In matrix form, $V = \left[\,\text{var}[T_{ij}]\,\right] = [m_{ij}^{(2)}] - [(m_{ij})^2] = M^{(2)} - M_{sq}$.

Example 2.1: **Special case**

The simplification of expressions for M and $M^{(2)}$ when $G\boldsymbol{e} = \boldsymbol{e}$ has been demonstrated. In [4] it was shown that matrix $G_{eb} = [g_{ij}] = [I - P + \boldsymbol{e}\boldsymbol{e}_b{}^T]^{-1}$ has many desirable characteristics. In particular it was shown that for such a matrix

$$\pi_j = g_{bj}, \quad j = 1, 2, ..., m \tag{2.35}$$

and $\quad m_{ij} = (\delta_{ij} + g_{jj} - g_{ij})/g_{bj} = \begin{cases} 1/g_{bj}, & i = j, \\ (g_{jj} - g_{ij})/g_{bj}, & i \neq j. \end{cases}$ (2.36)

Thus, following one matrix inversion, (actually only the *b-th* row for the stationary distribution), one can find the stationary probabilities and the mean first passage times. The efficiency of such a procedure is clear in that the inaccuracies alluded to in [1] and [2] are reduced to a minimum with the requirement that only an accurate package to compute a single matrix inverse (of a matrix whose elements do not need to be computed in advance) is required.

We consider using the matrix $G = G_{eb}$ to also find expressions for $M_d^{(2)}$ and $M^{(2)}$.
Firstly, from Cor 2.3.2, $M = D - GD + EG_dD$ where
$D = diag(1/g_{b1}, 1/g_{b2}, ..., 1/g_{bm}) = [\delta_{ij}/g_{bj}], GD = [g_{ij}/g_{bj}], EG_dD = [g_{jj}/g_{bj}]$

Further $\Pi G = \boldsymbol{e\pi}^T G = \boldsymbol{ee_b^T} G^2 = [\sum_{k=1}^m g_{bk}g_{kj}]$ so that

$(\Pi G)_d == [\delta_{ij} \sum_{k=1}^m g_{bk}g_{kj}] = [\delta_{ij} g_{bj}^{(2)}] = diag(g_{b1}^{(2)}, g_{b2}^{(2)}, ..., g_{bm}^{(2)})$.

From Cor 2.4.3, $M_d^{(2)} = D + 2G_dD^2 - 2(\Pi G)_d D^2$ where
$G_dD^2 = [\delta_{ij} g_{jj}/g_{bj}^2]$, $\Pi G = \boldsymbol{ee_b^T} G^2 = [g_{bj}^{(2)}]$, $(\Pi G)_d D^2 = [\delta_{ij}\, g_{bj}^{(2)}/g_{bj}^2]$
implying that $m_{jj}^{(2)} = (1/g_{bj}) + 2(g_{jj}/g_{bj}^2) - 2(g_{bj}^{(2)}/g_{bj}^2) = [g_{bj} + 2(g_{jj} - g_{bj}^{(2)})]/g_{bj}^2$.
Thus, from Eqn.(2.35)

$$m_{jj}^{(2)} = m_{jj} + 2m_{jj}^2(g_{jj} - g_{bj}^{(2)}). \tag{2.37}$$

Note that the only terms that needed to compute $m_{jj}^{(2)}$ involve the elements of the j-th column of G and the *(b,j)*-th element of G^2. In fact the computation of $M_d^{(2)}$ is effected through knowledge of G and the elements of the b-th row of G^2.

Note that since $\boldsymbol{Ge = e}$, $\sum_{k=1}^m g_{ik} = g_{i\bullet} = 1$, and thus

$$\sum_{k=1}^m (g_{ik} - g_{jk})m_{kj} = [(g_{ij} - g_{jj})m_{jj}] + [\sum_{k \neq j}^m (g_{ik} - g_{jk})(g_{jj} - g_{kj})m_{jj}]$$
$$= (g_{ij} - g_{jj})m_{jj} + \sum_{k=1}^m (g_{ik} - g_{jk})(g_{jj} - g_{kj})m_{jj}$$
$$= m_{jj}[(g_{ij} - g_{jj}) + \sum_{k=1}^m (g_{ik}g_{jj} - g_{ik}g_{kj} - g_{jk}g_{jj} + g_{jk}g_{kj})$$
$$= m_{jj}[(g_{ij} - g_{jj} + (\sum_{k=1}^m g_{ik})g_{jj} - g_{ij}^{(2)} - (\sum_{k=1}^m g_{jk})g_{jj} + g_{jj}^{(2)}]$$
$$= m_{jj}[g_{ij} - g_{jj} + g_{jj} - g_{ij}^{(2)} - g_{jj} + g_{jj}^{(2)} +] = m_{jj}[g_{ij} - g_{jj} + g_{jj}^{(2)} - g_{ij}^{(2)}].$$

From Eqn.(2.26), for $i \neq j$, using Eqn.(2.36),

$$m_{ij}^{(2)} = 2m_{jj}[g_{jj}^{(2)} - g_{ij}^{(2)}] + m_{ij}\left\{\frac{m_{jj}^{(2)}}{m_{jj}} - 2\right\}. \tag{2.38}$$

Eqns.(2.37) and (2.38) give elemental expressions for $m_{ij}^{(2)}$ for all i, j. Further $m_{ij}m_{jj}^{(2)} / m_{jj} = m_{ij}[1+2m_{jj}(g_{jj} - g_{bj}^{(2)})]$ so that

$$m_{ij}^{(2)} = \begin{cases} m_{jj}[1+2m_{jj}(g_{jj} - g_{bj}^{(2)})], & i = j, \\ 2m_{jj}[g_{jj}^{(2)} - g_{ij}^{(2)} + m_{ij}(g_{jj} - g_{bj}^{(2)})] - m_{ij}, & i \neq j. \end{cases} \tag{2.39}$$

These explicit elemental forms for $m_{ij}^{(2)}$, using the simple g-inverse G_{eb} = $[I - P + \boldsymbol{e}\boldsymbol{e}_b{}^T]^{-1}$ are new.

Further, $\mathrm{var}[T_{ij}] = \mathrm{var}[T_{ij} \mid X_0 = i] = m_{ij}^{(2)} - (m_{ij})^2$, and Eqn. (2.39) yields

$$\mathrm{var}[T_{ij}] = \begin{cases} m_{jj}[1 - m_{jj} + 2m_{jj}(g_{jj} - g_{bj}^{(2)})], & i = j, \\ 2m_{jj}[g_{jj}^{(2)} - g_{ij}^{(2)} + m_{ij}(g_{jj} - g_{bj}^{(2)})] - m_{ij}(1 - m_{ij}), & i \neq j. \end{cases} \tag{2.40}$$

We explore an application in Example 4.2. ❑

3 Variances of the mixing times

In [3] the "mixing matrix" was defined as $L = [\, l_{ij}] = [m_{ij}\pi_j\,]$ so that

$$L = M\Pi_d = MD^{-1} = M(M_d)^{-1}. \tag{3.1}$$

Note from Theorem 2.1, since $E\Pi_d = \Pi$,

$$(I - P)L = \Pi - P. \tag{3.2}$$

Theorem 3.1. *Let G be any g-inverse of I – P, then*

$$L = G\Pi - E(G\Pi)_d + I - G + EG_d, \tag{3.3}$$

If $G\boldsymbol{e} = g\boldsymbol{e}$ *then*

$$L = I - G + EG_d. \tag{3.4}$$

Further if $G = [g_{ij}]$ then, for all i, j,

$$l_{ij} = g_{jj} - g_{ij} + (g_{i\bullet} - g_{j\bullet})\pi_j + \delta_{ij}, \tag{3.5}$$

and if $G\boldsymbol{e} = g\boldsymbol{e}$ (*i.e.* $g_{i\bullet} = g$, *for all* i),

$$l_{ij} = g_{jj} - g_{ij} + \delta_{ij}. \tag{3.6}$$

Proof: Eqn.(3.3) follows from Eqn.(2.13) while Eqn.(3.4) follows from Eqn.(2.5). The elemental expressions, Eqns.(3.5) and (3.6), follow from Corollary 2.3.3. ❑

Since $\eta_i = \sum_{j=1}^{m} m_{ij}\pi_j$, if $\boldsymbol{\eta}^T = (\eta_1, \eta_2, ..., \eta_m)$, $\boldsymbol{\eta} = L\boldsymbol{e}$. Thus, from (3.2), since $\Pi\boldsymbol{e}$ = $\boldsymbol{e}\boldsymbol{\pi}^T\boldsymbol{e} = \boldsymbol{e}$ and $P\boldsymbol{e} = \boldsymbol{e}$,

$$(I - P)\boldsymbol{\eta} = \boldsymbol{0} \tag{3.7}$$

The irreducibility of the finite state Markov chain with transition matrix P implies ([10, Theorems 6.1.5 and 6.1.6]) that $\boldsymbol{\eta}$ is the right eigenvector corresponding to the eigenvalue $\lambda = 1$ and thus $\boldsymbol{\eta} = \eta\boldsymbol{e}$, for some η. Thus $\eta_i = \eta$, for all $i = 1, 2, ..., m$.

The following theorem summarises and extends the results given in [3] for expressions for η.

Theorem 3.2. *If $G = [g_{ij}]$ is any g-inverse of $I - P$,*

$$\boldsymbol{\eta} = [I - E(G\Pi)_d + EG_d]\boldsymbol{e} = [1 - tr(G\Pi) + tr(G)]\boldsymbol{e} = \eta\boldsymbol{e}. \tag{3.8}$$

Further if $g_{j\bullet} = \sum_{k=1}^{m} g_{jk}$, then

$$\eta = 1 + \sum_{j=1}^{m} (g_{jj} - g_{j\bullet}\pi_j). \tag{3.9}$$

Further, if $G\boldsymbol{e} = g\boldsymbol{e}$, $\qquad \eta = 1 - g + tr(G).$ (3.10)

In particular, $\qquad \eta = tr(Z) = 1 + tr(T).$ (3.11)

Proof: Eqn.(3.8) follows from (3.3) and $\boldsymbol{\eta} = L\boldsymbol{e}$ observing $E = \boldsymbol{e}\boldsymbol{e}^T$, $\boldsymbol{e}^T G_d \boldsymbol{e} = tr(G)$ and similary for $tr(G\Pi)$. Eqn.(3.9) is obtained by summing the expressions given by (3.5). Eqn.(3.10) follows by noting that under the condition $G\boldsymbol{e} = g\boldsymbol{e}$, $G\Pi = g\Pi$, and $tr(G\Pi) = g$. For Eqns.(3.11), when $G = Z$, $g = 1$, and when $G = T$, $g = 0$. ❑

We now wish to extend these results by exploring expressions for what we shall call the "second moment mixing matrix" defined as $L^{(2)} = [l_{ij}^{(2)}] = [m_{ij}^{(2)}\pi_j]$ so that

$$L^{(2)} = M^{(2)}\Pi_d = M^{(2)}D^{-1}. \tag{3.12}$$

Theorem 3.3. *If G is any g-inverse of $I - P$, then*

$$L^{(2)} = 2[GL - E(GL)_d] + [I - G + EG_d][L_d^{(2)} + I] - L, \tag{3.13}$$

$$= 2[GL - E(GL)_d] + 2[I - G + EG_d](\Pi L)_d D - L, \tag{3.14}$$

with

$$L_d^{(2)} = 2(\Pi L)_d D - I. \tag{3.15}$$

If $\boldsymbol{G}e = \boldsymbol{g}e$ then

$$L^{(2)} = 2[GL - E(GL)_d] + LL_d^{(2)}. \tag{3.16}$$

Proof: Eqns.(3.13) and (3.14) follow from Eqns.(2.14) and (2.15) observing that $M_d^{(2)}D^{-1} = L_d^{(2)}$ and $M = LD$ implying $E(GM)_d\Pi_d = E(GL)_d D\Pi_d = E(GL)_d$. Further $(\Pi M)_d = (\Pi L)_d D$. Eqn.(3.15) follows from Eqn.(2.8), and Eqn.(3.16) follows from Eqns. (3.13) or (3.14) using Eqn.(3.4) and $\Pi D = E$. ❑

Since $\eta_i^{(2)} = \sum_{j=1}^{m} m_{ij}^{(2)}\pi_j,$ we define $\boldsymbol{\eta}^{(2)T} = (\eta_1^{(2)}, \eta_2^{(2)}, \ldots, \eta_m^{(2)})$ and observe that $\boldsymbol{\eta}^{(2)} = L^{(2)}\boldsymbol{e}.$

Theorem 3.4. *If G is any g-inverse of $I - P$, then*

$$\boldsymbol{\eta}^{(2)} = 2[\eta G\boldsymbol{e} - tr(GL)\boldsymbol{e}] + 2[I - G + EG_d]\boldsymbol{\alpha} - \eta\boldsymbol{e}. \tag{3.17}$$

Further, if $G\boldsymbol{e} = g\boldsymbol{e}$,

$$\boldsymbol{\eta}^{(2)} = [2tr(G^2) - 3tr(G) - (1-2g)(1-g)]\boldsymbol{e} + 2L\boldsymbol{\alpha}. \tag{3.18}$$

In particular, when $g = 1$,

$$\boldsymbol{\eta}^{(2)} = [2tr(G^2) - 3tr(G)]\boldsymbol{e} + 2L\boldsymbol{\alpha}. \tag{3.19}$$

Also

$$\boldsymbol{\eta}^{(2)} = [2tr(Z^2) - 3tr(Z)]\boldsymbol{e} + 2L\boldsymbol{\alpha} = [2tr(T^2) - 3tr(T) - 1]\boldsymbol{e} + 2L\boldsymbol{\alpha}. \tag{3.20}$$

Proof: Since $\boldsymbol{\eta}^{(2)} = L^{(2)}\boldsymbol{e},$ Eqn.(3.17) follows from Eqn.(3.14) utilizing the results that $GL\boldsymbol{e} = G\boldsymbol{\eta} = \eta G\boldsymbol{e}$; $E(GL)_d\boldsymbol{e} = \boldsymbol{e}\boldsymbol{e}^T(GL)_d\boldsymbol{e} = \boldsymbol{e}\, tr(GL)$; $\boldsymbol{\alpha} = (\Pi M)_d\boldsymbol{e} = (\Pi L)_d D\boldsymbol{e}$ and $L\boldsymbol{e} = \eta\boldsymbol{e}$.
Under the condition $G\boldsymbol{e} = g\boldsymbol{e}$, from Eqn.(3.4), $L = I - G + EG_d$.
Further $GL = G[I - G + EG_d] = G - G^2 + GEG_d = G - G^2 + gEG_d$.
Since $tr(EG_d) = tr(G)$, $tr(GL) = (1 + g)tr(G) - tr(G^2)$ and Eqn.(3.18) follows from Eqn.(3.17) using Eqn.(3.10), and Eqn.(3.19) from (3.18) when $g = 1$.
Eqn.(3.20) follows upon simplification with $g = 1$ when $G = Z$ and $g = 0$ when $G = T$.❑

As discussed in Section 1, ν_i, the variance of the mixing time when the Markov chain starts in state i, is given by $\eta_i^{(2)} - (\eta_i)^2 = \eta_i^{(2)} - \eta^2$. Thus if $\boldsymbol{\nu}^T = (\nu_1, \nu_2, \ldots, \nu_m)$ we have that $\boldsymbol{\nu} = \boldsymbol{\eta}^{(2)} - \eta^2\boldsymbol{e}$. The following key result now follows using the expressions for $\boldsymbol{\eta}^{(2)}$ and $\boldsymbol{\eta}$ given by Theorems 3.2 and 3.4.

Theorem 3.5. *If G is any g-inverse of $I - P$, then*

$$\boldsymbol{\nu} = 2[I - G + EG_d]\boldsymbol{\alpha} + [2\eta G - 2tr(GL) - \eta - \eta^2]\boldsymbol{e}. \tag{3.21}$$

Further, if $G\boldsymbol{e} = g\boldsymbol{e}$,

$$\boldsymbol{\nu} = 2L\boldsymbol{\alpha} + [2tr(G^2) - (tr(G))^2 - (5-2g)tr(G) - (1-g)(2-3g)]\boldsymbol{e}. \tag{3.22}$$

In particular, if $g = 1$,

$$\boldsymbol{\nu} = 2L\boldsymbol{\alpha} + [2tr(G^2) - (tr(G))^2 - 3tr(G)]\boldsymbol{e}. \tag{3.23}$$

Also,

$$\boldsymbol{\nu} = 2L\boldsymbol{\alpha} + [2tr(Z^2) - (tr(Z))^2 - 3tr(Z)]\boldsymbol{e}, \tag{3.24}$$

$$\boldsymbol{\nu} = 2L\boldsymbol{\alpha} + [2tr(T^2) - (tr(T))^2 - 5tr(T) - 2]\boldsymbol{e}. \tag{3.25}$$

Proof: Eqn.(3.21) follows from (3.17), Eqns.(3.22) and (3.23) from (3.18) noting, from (3.10), that $\eta = 1 - g + tr(G)$, while (3.24) follows from (3.23) with $G = Z$ and (3.25) from (3.22) with $G = T$ and $g = 0$ respectively. ❑

A key observation from Theorem 3.5 is that $v_i = \text{var}\,[T \mid X_0 = i] = v$, constant for all i, if and only if $L\boldsymbol{\alpha} = l\,\boldsymbol{e}$ for some constant l. Since $L\boldsymbol{e} = \eta\boldsymbol{e}$, a sufficient condition is that $\boldsymbol{\alpha} = \alpha\boldsymbol{e}$ for some α.

Since, from Eqn.(2.29), $\boldsymbol{\alpha} = (\boldsymbol{\pi}^T G\boldsymbol{e})\boldsymbol{e} - (GE)_d\boldsymbol{e} + \boldsymbol{e} - (\Pi G)_d D\boldsymbol{e} + G_d D\boldsymbol{e}$, this requires

$[\alpha - \boldsymbol{\pi}^T G\boldsymbol{e} - 1]\boldsymbol{e} = [G_d D - (GE)_d - (\Pi G)_d D]\boldsymbol{e}$

$\Leftrightarrow G_d D - (GE)_d - (\Pi G)_d D = kI$ for some k.

4 Special cases

Example 4.1. *Two-state Markov chains*

Let $P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$, $(0 \le a \le 1,\ 0 \le b \le 1)$, be the transition matrix of a two-state Markov chain with state space $S = \{1, 2\}$. Let $d = 1 - a - b$ so that $1 - d = a + b$.

If $-1 \le d < 1$, the Markov chain is irreducible with a unique stationary distribution given by

$$\pi_1 = \frac{b}{1-d},\ \pi_2 = \frac{a}{1-d}.$$

If $-1 < d < 1$, the Markov chain is regular and this stationary distribution is in fact the limiting distribution. If $d = -1$ the Markov chain is irreducible periodic, period 2.

Thus for $-1 \le d < 1$, $\Pi = \dfrac{1}{1-d}\begin{bmatrix} b & a \\ b & a \end{bmatrix}$.

Every g-inverse of $I - P$ can be expressed as $G(\boldsymbol{t}, \boldsymbol{u}) + \boldsymbol{e}\boldsymbol{f}^T + \boldsymbol{g}\boldsymbol{\pi}^T$ (see [10], [11]) where $G(\boldsymbol{t}, \boldsymbol{u}) = [I - P + \boldsymbol{t}\boldsymbol{u}^T]^{-1}$, provided $\boldsymbol{\pi}^T\boldsymbol{t} \neq 0$, $\boldsymbol{u}^T\boldsymbol{e} \neq 0$.

For the above two-state Markov chain

$$G \equiv G(\boldsymbol{t},\boldsymbol{u}) = \frac{1}{(bt_1 + at_2)(u_1 + u_2)}\begin{bmatrix} b + t_2u_2 & a - t_1u_2 \\ b - t_2u_1 & a + t_1u_1 \end{bmatrix}.$$

Provided $-1 \le d < 1$, the fundamental matrix $Z = G(\boldsymbol{e},\boldsymbol{\pi})$, so that taking $u_1 = b/(1-d)$, $u_2 = a/(1-d)$,

$$Z = [I - P + \Pi]^{-1} = \frac{1}{1-d}\begin{bmatrix} b + \dfrac{a}{1-d} & a - \dfrac{a}{1-d} \\ b - \dfrac{b}{1-d} & a + \dfrac{b}{1-d} \end{bmatrix}, \text{([13]).}$$

The group inverse $T = (I - \Pi)G(I - \Pi) = Z - \Pi$ now follows, as $T = \frac{1}{(1-d)^2}\begin{bmatrix} a & -a \\ -b & b \end{bmatrix}$.

The mean first passage time matrix, $M = [G\Pi - E(G\Pi)_d + I - G + EG_d]D$, where

$$G\Pi - E(G\Pi)_d = \frac{1}{(bt_1 + at_2)(1-d)}\begin{bmatrix} 0 & (t_2 - t_1)a \\ (t_1 - t_2)b & 0 \end{bmatrix} \text{ (= 0 when } t_1 = t_2\text{, as for the case } G(\boldsymbol{e},\boldsymbol{u}))$$

$$I - G + EG_d = \frac{1}{bt_1 + at_2}\begin{bmatrix} bt_1 + at_2 & t_1 \\ t_2 & bt_1 + at_2 \end{bmatrix} \left(= \frac{1}{1-d}\begin{bmatrix} 1-d & 1 \\ 1 & 1-d \end{bmatrix} \text{ when } t_1 = t_2\right)$$

leading to $M = \begin{bmatrix} \frac{1-d}{b} & \frac{1}{a} \\ \frac{1}{b} & \frac{1-d}{a} \end{bmatrix}$, ([10, p. 135], [13, p. 94]).

Now $M_d = D = \begin{bmatrix} \frac{1-d}{b} & 0 \\ 0 & \frac{1-d}{a} \end{bmatrix}$ and $\Pi M = \begin{bmatrix} 1 + \frac{a}{b(1-d)} & 1 + \frac{b}{a(1-d)} \\ 1 + \frac{a}{b(1-d)} & 1 + \frac{b}{a(1-d)} \end{bmatrix}$ so that

$$M_d^{(2)} = 2D(\Pi M)_d - D = \begin{bmatrix} \frac{1-d}{b} + \frac{2a}{b^2} & 0 \\ 0 & \frac{1-d}{a} + \frac{2b}{a^2} \end{bmatrix}$$

$$M^{(2)} = 2[GM - E(GM)_d] + [I - G + EG_d][M_d^{(2)} + D] - M = \begin{bmatrix} \frac{2a}{b^2} + \frac{(1-d)}{b} & \frac{2}{a^2} - \frac{1}{a} \\ \frac{2}{b^2} - \frac{1}{b} & \frac{2b}{a^2} + \frac{(1-d)}{a} \end{bmatrix}.$$

(Simplification takes place using $ad + b = (1-d)(1-a)$ and $bd + a = (1-d)(1-b)$.)

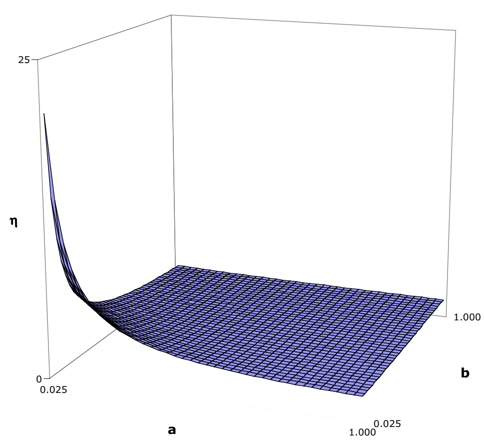
Since $M_{sq} = \begin{bmatrix} \frac{(1-d)^2}{b^2} & \frac{1}{a^2} \\ \frac{1}{b^2} & \frac{(1-d)^2}{a^2} \end{bmatrix}$, we deduce that

$$[\operatorname{var} T_{ij}] = M^{(2)} - M_{sq} = \begin{bmatrix} \frac{a(1+d)}{b^2} & \frac{1-a}{a^2} \\ \frac{1-b}{b^2} & \frac{b(1+d)}{a^2} \end{bmatrix}, \text{ ([10, p. 135], [13, p. 94]).}$$

The mixing matrix $L = G\Pi - E(G\Pi)_d + I - G + EG_d = \begin{bmatrix} 1 & \dfrac{1}{1-d} \\ \dfrac{1}{1-d} & 1 \end{bmatrix}$,

so that the expected time to mixing, $\eta = L\boldsymbol{e} = 1 + \dfrac{1}{1-d} = 1 + \dfrac{1}{a+b}$.

As shown in [3], for all two-state irreducible Markov chains, $\eta \geq 1.5$ with the minimum value of $\eta = 1.5$ occurring when $d = -1$, ($a = 1$, $b = 1$) in which case the Markov chain is periodic, period *2*. Arbitrarily large values of η occur as $d \to 1$, (when both $a \to 0$ and $b \to 0$), when the chain is approaching the situation when the chain is close to being reducible with both states absorbing. Graph 1 displays the expected time to mixing for a, b taking values 0.025 to 1.000 in steps of 0.025.



Graph 1: Expected time to mixing

To find the second moment mixing matrix first observe that

$$L_d^{(2)} = 2(\Pi L)_d D - I = 2(\Pi M)_d - I = \begin{bmatrix} 1 + \dfrac{2a}{b(1-d)} & 0 \\ 0 & 1 + \dfrac{2b}{a(1-d)} \end{bmatrix}, \text{ and hence}$$

$$L^{(2)} = 2[GL - E(GL)_d] + [I - G + EG_d][L_d^{(2)} + I] - L = \begin{bmatrix} 1 + \dfrac{2a}{b(1-d)} & \dfrac{2-a}{a(1-d)} \\ \dfrac{2-b}{b(1-d)} & 1 + \dfrac{2b}{a(1-d)} \end{bmatrix}.$$

The second moments of the mixing times, $\boldsymbol{\eta}^{(2)} = L^{(2)}\boldsymbol{e} = \begin{bmatrix} 1 + \dfrac{2a}{b(1-d)} + \dfrac{2-a}{a(1-d)} \\ 1 + \dfrac{2b}{a(1-d)} + \dfrac{2-b}{b(1-d)} \end{bmatrix}$.

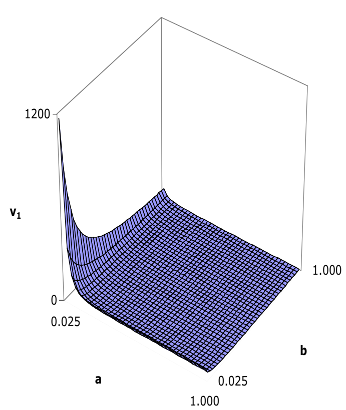
Observe also that $\boldsymbol{\alpha} = (\Pi M)_d \boldsymbol{e} = \begin{bmatrix} 1 + \dfrac{a}{b(1-d)} \\ 1 + \dfrac{b}{a(1-d)} \end{bmatrix}$.

The variances of the mixing times, starting in state i, are given by ν_i where

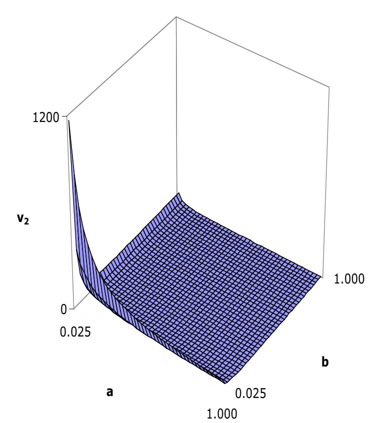
$$\boldsymbol{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} = \boldsymbol{\eta}^{(2)} - \eta^2 \boldsymbol{e} = \begin{bmatrix} \dfrac{2a}{b(1-d)} + \dfrac{2}{a(1-d)} - \dfrac{3}{1-d} - \left(\dfrac{1}{1-d}\right)^2 \\ \dfrac{2b}{a(1-d)} + \dfrac{2}{b(1-d)} - \dfrac{3}{1-d} - \left(\dfrac{1}{1-d}\right)^2 \end{bmatrix}$$

$$= \frac{1}{ab(1-d)^2} \begin{bmatrix} (2a^2 + 2b - 3ab)(a+b) - ab \\ (2b^2 + 2a - 3ab)(a+b) - ab \end{bmatrix}.$$

We graph the values of ν_1 (Graph 2) and ν_2 (Graph 3) for all values of a and b between 0.025 and 1.000 in steps of 0.025. (Observe that $\nu_1 \to \infty$ and $\nu_2 \to \infty$ as both $a \to 0$ and $b \to 0$.)



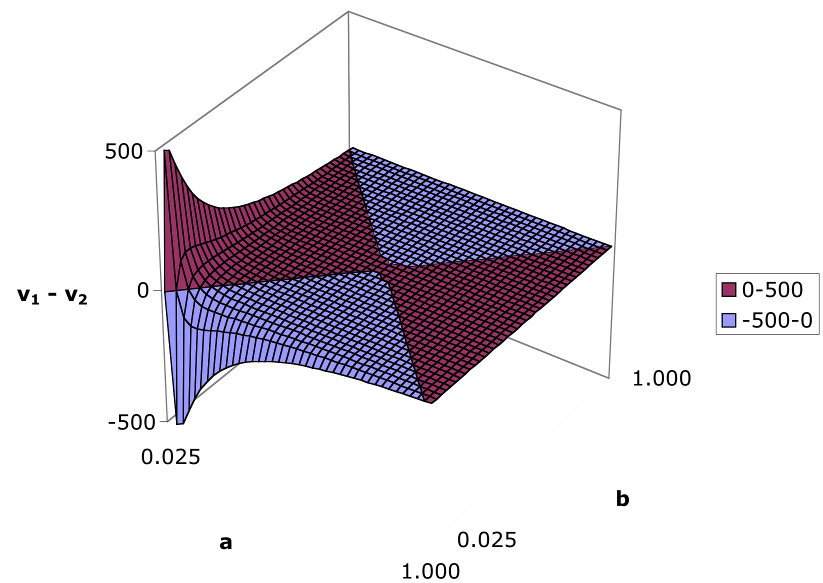
Graph 2: Variance$_1$



Graph 3: Variance$_2$

It is easy to show that $\nu_1 < \nu_2 \Leftrightarrow d(a-b) > 0 \Leftrightarrow$ either *(i)* $a > b$ and $a + b < 1$ or *(ii)* $a < b$ and $a + b > 1$.
The lines $a = b$ and $a + b = 1$ partition the parameter space (a, b) into regions where $\nu_1 = \nu_2$, $\nu_1 < \nu_2$, and $\nu_1 > \nu_2$, as illustrated on Graph 4.

The significance of this is that if we are given the transition probabilities, or the transition matrix $P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$, of the Markov chain and we wish to minimise the variance of the time to mixing one would choose state 1 as the starting



Graph 4: Variance$_1$ – Variance$_2$

state to achieve $\nu_1 < \nu_2$. This is equivalent to choosing state 1 if either $p_{21} < p_{11} < p_{22}$ or $p_{22} < p_{11} < p_{21}$ (or, equivalently, if $p_{21} < p_{12} < p_{22}$ or $p_{22} < p_{12} < p_{21}$) otherwise choose state 2.

If $a = b$ (i.e. $p_{12} = p_{21}$ or $p_{11} = p_{22}$) or $a + b = 1$ (i.e. $p_{11} = p_{21}$ or $p_{12} = p_{22}$) the choice of the starting state is immaterial. The later case is equivalent to independent trials.

Note that as $a \to 0$ and $b \to 0$ both of the variances ν_1 and ν_2 $\to \infty$ but one needs to consider how the limit is approached since the difference $\nu_1 - \nu_2$ can approach $\to \infty$ or $\to -\infty$ or be equal (and both $\to \infty$). In these situations the MC is tending to an absorbing MC and close to being reducible. The expected time to mixing in such a chain can also be arbitrarily large, as considered in [3].

Note that the minimum variances of the mixing times ($\nu_1 = 0.25$, $\nu_2 = 0.25$) occur when a = b = 1, (with a minimum of 0.25) with the expected mixing time also at a minimum of $\eta = 1.5$.

Example 4.2. *Three-state Markov chains*

Let $P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 1-a_2-a_3 & a_2 & a_3 \\ b_1 & 1-b_1-b_3 & b_3 \\ c_1 & c_2 & 1-c_1-c_2 \end{bmatrix}$ be the transition matrix of a Markov chain with state space $S = \{1, 2, 3\}$. Note that $0 < a_2 + a_3 \le 1$, $0 < b_1 + b_3 \le 1$ and $0 < c_1 + c_2 \le 1$.

Let $\Delta_1 \equiv b_3c_1 + b_1c_2 + b_1c_1$, $\Delta_2 \equiv c_1a_2 + c_2a_3 + c_2a_2$, $\Delta_3 \equiv a_2b_3 + a_3b_1 + a_3b_3$,
$\Delta \equiv \Delta_1 + \Delta_2 + \Delta_3$,

The Markov chain, with the above transition matrix, is irreducible (and hence a stationary distribution exists) if and only if $\Delta_1 > 0$, $\Delta_2 > 0$, $\Delta_3 > 0$.

It is easily shown that the stationary probability vector is $(\pi_1, \pi_2, \pi_3) = \frac{1}{\Delta}(\Delta_1, \Delta_2, \Delta_3)$,

so that $\Pi = \frac{1}{\Delta}\begin{bmatrix} \Delta_1 & \Delta_2 & \Delta_3 \\ \Delta_1 & \Delta_2 & \Delta_3 \\ \Delta_1 & \Delta_2 & \Delta_3 \end{bmatrix}$ so that $D = \Pi_d^{-1} = \Delta\begin{bmatrix} 1/\Delta_1 & 0 & 0 \\ 0 & 1/\Delta_2 & 0 \\ 0 & 0 & 1/\Delta_3 \end{bmatrix}$.

The matrix $G(\boldsymbol{t}, \boldsymbol{u}) = [I - P + \boldsymbol{t}\boldsymbol{u}^T\,]^{-1}$,where $\boldsymbol{\pi}^T\boldsymbol{t} \neq 0$, $\boldsymbol{u}^T\boldsymbol{e} \neq 0$, can be taken as a g-inverse of $I - P$. It can be shown that any g-inverse of $I - P$ can be can be expressed as $G(\boldsymbol{t}, \boldsymbol{u}) + \boldsymbol{e}\boldsymbol{f}^T + \boldsymbol{g}\boldsymbol{\pi}^T$. (see [9], [10].)

In [3] we showed that for the above three-state Markov chain that if
$\Delta(\boldsymbol{t}, \boldsymbol{u}) = (\Delta_1 t_1 + \Delta_2 t_2 + \Delta_3 t_3)(u_1 + u_2 + u_3)$, then

$$G(\boldsymbol{t},\boldsymbol{u}) = \frac{1}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} \Delta_1 & \Delta_2 & \Delta_3 \\ \Delta_1 & \Delta_2 & \Delta_3 \\ \Delta_1 & \Delta_2 & \Delta_3 \end{bmatrix} +$$

$$\frac{t_1u_1}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} 0 & 0 & 0 \\ 0 & c_1+c_2 & b_3 \\ 0 & c_2 & b_1+b_3 \end{bmatrix} + \frac{t_1u_2}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} 0 & -(c_1+c_2) & -b_3 \\ 0 & 0 & 0 \\ 0 & -c_1 & b_1 \end{bmatrix} + \frac{t_1u_3}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} 0 & -c_2 & -(b_1+b_3) \\ 0 & c_1 & -b_1 \\ 0 & 0 & 0 \end{bmatrix} +$$

$$\frac{t_2u_1}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} 0 & 0 & 0 \\ -(c_1+c_2) & 0 & -a_3 \\ -c_2 & 0 & a_2 \end{bmatrix} + \frac{t_2u_2}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} c_1+c_2 & 0 & a_3 \\ 0 & 0 & 0 \\ c_1 & 0 & a_2+a_3 \end{bmatrix} + \frac{t_2u_3}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} c_2 & 0 & -a_2 \\ -c_1 & 0 & -(a_2+a_3) \\ 0 & 0 & 0 \end{bmatrix} +$$

$$\frac{t_3u_1}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} 0 & 0 & 0 \\ -b_3 & a_3 & 0 \\ -(b_1+b_3) & -a_2 & 0 \end{bmatrix} + \frac{t_3u_2}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} b_3 & -a_3 & 0 \\ 0 & 0 & 0 \\ -b_1 & -(a_2+a_3) & 0 \end{bmatrix} + \frac{t_3u_3}{\Delta(\boldsymbol{t},\boldsymbol{u})}\begin{bmatrix} b_1+b_3 & a_2 & 0 \\ b_1 & a_2+a_3 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Define $\tau_{12} = a_3 + c_1 + c_2$, $\tau_{13} = a_2 + b_1 + b_3$, $\tau_{21} = b_3 + c_1 + c_2$, $\tau_{23} = b_1 + a_2 + a_3$, $\tau_{31} = c_2 + b_1 + b_3$, $\tau_{32} = c_1 + a_2 + a_3$, $\tau = a_2 + a_3 + b_1 + b_3 + c_1 + c_2$, so that $\tau = \tau_{12} + \tau_{13} = \tau_{21} + \tau_{23} = \tau_{31} + \tau_{32}$.

Now $\Delta(\boldsymbol{e}, \boldsymbol{u}) = \Delta(u_1 + u_2 + u_3)$, so that

$$G(e,\boldsymbol{u}) = \Pi + \frac{u_1}{\Delta}\begin{bmatrix} 0 & 0 & 0 \\ -\tau_{21} & \tau_{12} & b_3 - a_3 \\ -\tau_{31} & c_2 - a_2 & \tau_{13} \end{bmatrix} + \frac{u_2}{\Delta}\begin{bmatrix} \tau_{21} & -\tau_{12} & a_3 - b_3 \\ 0 & 0 & 0 \\ c_1 - b_1 & -\tau_{32} & \tau_{23} \end{bmatrix} + \frac{u_3}{\Delta}\begin{bmatrix} \tau_{31} & a_2 - c_2 & -\tau_{13} \\ b_1 - c_1 & \tau_{32} & -\tau_{23} \\ 0 & 0 & 0 \end{bmatrix}$$

Observe that $G(e,\boldsymbol{u})\boldsymbol{e} = \boldsymbol{e}$ implying that $g = 1$.

Note that explicit expressions for $Z = G(\boldsymbol{e}, \boldsymbol{\pi})$ and $T = Z - \Pi$ follow. In particular

$$T = \frac{\Delta_1}{\Delta^2}\begin{bmatrix} 0 & 0 & 0 \\ -\tau_{21} & \tau_{12} & b_3 - a_3 \\ -\tau_{31} & c_2 - a_2 & \tau_{13} \end{bmatrix} + \frac{\Delta_2}{\Delta^2}\begin{bmatrix} \tau_{21} & -\tau_{12} & a_3 - b_3 \\ 0 & 0 & 0 \\ c_1 - b_1 & \tau_{32} & \tau_{23} \end{bmatrix} + \frac{\Delta_3}{\Delta^2}\begin{bmatrix} \tau_{31} & a_2 - c_2 & -\tau_{13} \\ b_1 - c_1 & \tau_{32} & -\tau_{23} \\ 0 & 0 & 0 \end{bmatrix}.$$

One of the simplest g-inverses of $I - P$ that we can take is

$$G \equiv G(\boldsymbol{e},\boldsymbol{e}_1) = \left[g_{ij}\right] = \frac{1}{\Delta}\begin{bmatrix} \Delta_1 & \Delta_2 & \Delta_3 \\ \Delta_1 - \tau_{21} & \Delta_2 + \tau_{12} & \Delta_3 + \tau_{21} - \tau_{12} \\ \Delta_1 - \tau_{31} & \Delta_2 + \tau_{31} - \tau_{13} & \Delta_3 + \tau_{13} \end{bmatrix}.$$

Using G, as defined above, the mean first passage time matrix can be found using Eqn. (2.36) as

$$m_{ij} = \frac{\delta_{ij} + g_{jj} - g_{ij}}{g_{1j}} = \begin{cases} \dfrac{\Delta}{\Delta_j}, & i = j, \\[2ex] \dfrac{\tau_{ij}}{\Delta_j}, & i \neq j; \end{cases}$$

implying

$$M = \begin{bmatrix} \dfrac{\Delta}{\Delta_1} & \dfrac{\tau_{12}}{\Delta_2} & \dfrac{\tau_{13}}{\Delta_3} \\[2ex] \dfrac{\tau_{21}}{\Delta_1} & \dfrac{\Delta}{\Delta_2} & \dfrac{\tau_{23}}{\Delta_3} \\[2ex] \dfrac{\tau_{31}}{\Delta_1} & \dfrac{\tau_{32}}{\Delta_2} & \dfrac{\Delta}{\Delta_3} \end{bmatrix}, \text{ (as also derived in [3]).}$$

Define $\kappa_1 = \Delta_2\tau_{21} + \Delta_3\tau_{31}$, $\kappa_2 = \Delta_1\tau_{12} + \Delta_3\tau_{32}$, $\kappa_3 = \Delta_1\tau_{13} + \Delta_2\tau_{23}$ and note that $\kappa_1 + \kappa_2 + \kappa_3 = \Delta\tau$. Also let $\rho_j \equiv \kappa_j/\Delta_j$.

We compute $G^2 = \left[g_{ij}^{(2)}\right] = \left[\sum_{k=1}^{3} g_{ik}g_{kj}\right]$ yielding, for the first row, after simplification,

$$g_{11}^{(2)} = \frac{\Delta_1}{\Delta} - \frac{\Delta_2\tau_{21} + \Delta_3\tau_{31}}{\Delta^2},$$

$$g_{12}^{(2)} = \frac{\Delta_2}{\Delta} + \frac{\Delta_2\tau_{12} + \Delta_3(\tau_{31} - \tau_{13})}{\Delta^2},\ g_{13}^{(2)} = \frac{\Delta_3}{\Delta} + \frac{\Delta_2(\tau_{21} - \tau_{12}) + \Delta_3\tau_{13}}{\Delta^2}.$$

From these terms, the expressions for the m_{ij}, and Eqn. (2.37), i.e.

$$m_{jj}^{(2)} = m_{jj} + 2m_{jj}^2(g_{jj} - g_{1j}^{(2)}),$$

we immediately obtain expressions for the $m_{jj}^{(2)}$ which, after simplification, yield

$$m_{jj}^{(2)} = \frac{\Delta}{\Delta_j} + \frac{2\kappa_j}{\Delta_j^2} = \frac{\Delta + 2\rho_j}{\Delta_j} = m_{jj} + \frac{2\rho_j}{\Delta_j}, j = 1, 2, 3. \tag{4.1}$$

We now compute the remaining $g_{ij}^{(2)}$ terms (re-expressing and simplifying where necessary):

$$g_{21}^{(2)} = \frac{\Delta_1}{\Delta} - \frac{\Delta_2\tau_{21} + \Delta_3\tau_{31}}{\Delta^2} + \frac{[\tau_{12}\tau_{31} - \tau_{12}\tau_{21} - \tau_{21}\tau_{31}]}{\Delta^2},$$

$$g_{22}^{(2)} = \frac{\Delta_2}{\Delta} + \frac{\Delta_2\tau_{12} + \Delta_3(\tau_{31} - \tau_{13})}{\Delta^2} + \frac{[\tau_{12}\tau_{32} + \tau_{21}\tau_{31} - \tau_{13}\tau_{21}]}{\Delta^2},$$

$$g_{23}^{(2)} = \frac{\Delta_3}{\Delta} + \frac{\Delta_2(\tau_{21} - \tau_{12}) + \Delta_3\tau_{13}}{\Delta^2} + \frac{[\tau_{21}\tau_{13} - \tau_{12}\tau_{23}]}{\Delta^2},$$

$$g_{31}^{(2)} = \frac{\Delta_1}{\Delta} - \frac{\Delta_2\tau_{21} + \Delta_3\tau_{31}}{\Delta^2} + \frac{[\tau_{13}\tau_{21} - \tau_{21}\tau_{31} - \tau_{13}\tau_{31}]}{\Delta^2},$$

$$g_{32}^{(2)} = \frac{\Delta_2}{\Delta} + \frac{\Delta_2\tau_{12} + \Delta_3(\tau_{31} - \tau_{13})}{\Delta^2} + \frac{\tau_{12}\tau_{31} - \tau_{13}\tau_{32}}{\Delta^2},$$

$$g_{33}^{(2)} = \frac{\Delta_3}{\Delta} + \frac{\Delta_2(\tau_{21} - \tau_{12}) + \Delta_3\tau_{13}}{\Delta^2} + \frac{[\tau_{21}\tau_{31} + \tau_{13}\tau_{23} - \tau_{12}\tau_{31}]}{\Delta^2}.$$

From these results it can be shown that

$$g_{jj}^{(2)} - g_{ij}^{(2)} = \frac{[\tau_{ij}\tau_{ji} + \tau_{kj}(\tau_{ij} - \tau_{ji})]}{\Delta^2}, \quad i \neq j \neq k \in \{1, 2, 3\}.$$

Substitution in Eqn.(2.38), yields for $i \neq j$,

$$m_{ij}^{(2)} = \frac{2[\tau_{ij}\tau_{ji} + \tau_{kj}(\tau_{ij} - \tau_{ji}) + \tau_{ij}\rho_j] - \tau_{ij}\Delta}{\Delta\Delta_j}. \tag{4.2}$$

The explicit elemental expressions for the mean first passage times in a general three-state Markov chain given by Eqns. (4.1) and (4.2) are new results.

Alternatively, using G, as above, $\{(I - \Pi)G\}_d = \frac{1}{\Delta^2}\begin{bmatrix} \kappa_1 & 0 & 0 \\ 0 & \kappa_2 & 0 \\ 0 & 0 & \kappa_3 \end{bmatrix}$, and Eqn.(2.11),

yields, after simplification,

$$M_d^{(2)} = \begin{bmatrix} (\Delta + 2\rho_1)/\Delta_1 & 0 & 0 \\ 0 & (\Delta + 2\rho_2)/\Delta_2 & 0 \\ 0 & 0 & (\Delta + 2\rho_3)/\Delta_3 \end{bmatrix}.$$

Further from Eqn.(2.22), $M^{(2)} = 2[GM - E(GM)_d] + MD^{-1}M_d^{(2)}$ where the diagonal elements of the first expression $2[GM - E(GM)_d]$ are all 0, and the diagonal elements of the second expression are $M_dD^{-1}M_d^{(2)} = M_d^{(2)}$.

By evaluating each element of GM it can be shown, upon simplification:

$$(GM)_{ij} - (GM)_{jj} = \begin{cases} [\tau_{ij}\tau_{ji} + \tau_{kj}(\tau_{ij} - \tau_{ji}) - \Delta\tau_{ij}] / \Delta\Delta_j, & i \neq j \neq k \in \{1,2,3\}, \\ 0, & i = j. \end{cases}$$

Further $(MD^{-1}M_d^{(2)})_{ij} = \begin{cases} \tau_{ij}(\Delta + 2\rho_j) / \Delta\Delta_j, & i \neq j, \\ (\Delta + 2\rho_j) / \Delta_j, & i = j, \end{cases}$ leading to Eqns.(4.1) and (4.2) above.

Expressions for the variances of each first passage time r.v. T_{ij} follow from elemental expressions of $V = [\,var(T_{ij})] = M^{(2)} - M_{sq}$.

$$V = [m_{ij}^{(2)} - m_{ij}^2] = \begin{cases} [2\tau_{ij}\tau_{ji} + 2\tau_{kj}(\tau_{ij} - \tau_{ji}) + 2\tau_{ij}\rho_j - \Delta\tau_{ij}] / \Delta\Delta_j - (\tau_{ij} / \Delta_j)^2, & i \neq j\ (\neq\ k), \\ ([\Delta + 2\rho_j] / \Delta_j)\ -\ (\Delta/\Delta_j)^2, & i = j. \end{cases}$$

The matrix of 'mixing times' are $L = \left[m_{ij}\pi_j \right] = \dfrac{1}{\Delta}\begin{bmatrix} \Delta & \tau_{12} & \tau_{13} \\ \tau_{21} & \Delta & \tau_{23} \\ \tau_{31} & \tau_{32} & \Delta \end{bmatrix}$.

The common row sums of L lead to the expected "time to mixing",

$$\eta = 1\ +\ (\tau/\Delta). \tag{4.3}$$

In Hunter [3] it was shown that for all three-state irreducible Markov chains, $\eta \geq 2$, with $\eta = 2$ achieved by "the minimal period 3" case (See Case 1 to follow.).

Elemental expressions for "second moments of mixing", $l_{ij}^{(2)} = m_{ij}^{(2)}\pi_j$, follow from Eqns.(4,1) and (4.2). These lead to the second moment of the mixing time starting in state i as $\eta_i^{(2)} = \sum_{j=1}^{m} m_{ij}^{(2)}\pi_j$. Derivation of simple forms for these elemental expressions, when evaluated directly from the stationary probabilities and the second first passage time moments, requires considerable effort. However, from Eqn.(3.18) of Theorem 3.4, with $\boldsymbol{\alpha} = (\Pi M)_d \boldsymbol{e}$, $\boldsymbol{\eta}^{(2)} = [2tr(G^2) - 3tr(G)]\boldsymbol{e} + 2L\boldsymbol{\alpha}$.

$$\boldsymbol{\alpha} = (\Pi M)_d \boldsymbol{e} = \begin{bmatrix} 1 + \dfrac{\kappa_1}{\Delta\Delta_1} & 0 & 0 \\ 0 & 1 + \dfrac{\kappa_2}{\Delta\Delta_2} & 0 \\ 0 & 0 & 1 + \dfrac{\kappa_3}{\Delta\Delta_3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 + \dfrac{\kappa_1}{\Delta\Delta_1} \\ 1 + \dfrac{\kappa_2}{\Delta\Delta_2} \\ 1 + \dfrac{\kappa_3}{\Delta\Delta_3} \end{bmatrix} = \boldsymbol{e} + \frac{1}{\Delta}\boldsymbol{\rho},$$

with $\boldsymbol{\rho}^T = (\rho_1, \rho_2, \rho_3)$ implying that $L\boldsymbol{\alpha} = L\boldsymbol{e} + \dfrac{1}{\Delta}L\boldsymbol{\rho} = \left(1 + \dfrac{\tau}{\Delta}\right)\boldsymbol{e} + \dfrac{1}{\Delta^2}\boldsymbol{\beta}$, where

$$\boldsymbol{\beta} = \begin{bmatrix} \Delta\rho_1 + \tau_{12}\rho_2 + \tau_{13}\rho_3 \\ \tau_{21}\rho_1 + \Delta\rho_2 + \tau_{23}\rho_3 \\ \tau_{31}\rho_1 + \tau_{32}\rho_2 + \Delta\rho_3 \end{bmatrix}.$$

Further

$$tr(G) = g_{11} + g_{22} + g_{33} = 1 + \frac{\tau}{\Delta} = \eta,$$

$$tr(G^2) = g_{11}^{(2)} + g_{22}^{(2)} + g_{33}^{(2)} = 1 + \frac{[\tau^2 - 2(\tau_{21}\tau_{13} + \tau_{12}\tau_{31} - \tau_{21}\tau_{31})]}{\Delta^2} = 1 - \frac{2}{\Delta} + \frac{\tau^2}{\Delta^2}.$$

so that

$$\boldsymbol{\eta}^{(2)} = [1 - \frac{\tau}{\Delta} - \frac{4}{\Delta} + \frac{2\tau^2}{\Delta^2}]\boldsymbol{e} + \frac{2}{\Delta^2}\boldsymbol{\beta}. \tag{4.4}$$

If the variance of the mixing time when the Markov chain starts in state i, is v_i, let the vector of the variances be $\boldsymbol{v}^T = (v_1, v_2, ..., v_m)$. Since $\mathbf{v} = \boldsymbol{\eta}^{(2)} - \eta^2\boldsymbol{e}$, from Eqns. (4.3) and (4.4),

$$\boldsymbol{v} = [\frac{\tau^2}{\Delta^2} - \frac{3\tau}{\Delta} - \frac{4}{\Delta}]\boldsymbol{e} + \frac{2}{\Delta^2}\boldsymbol{\beta}. \tag{4.5}$$

The expression given by Eqn.(4.5) shows that variability between the variances of the mixing times is explained by the elements of the vector $\boldsymbol{\beta}$, i.e. $\Delta\rho_1 + \tau_{12}\rho_2 + \tau_{13}\rho_3$ for state 1, $\tau_{21}\rho_1 + \Delta\rho_2 + \tau_{23}\rho_3$ for state 2, and $\tau_{31}\rho_1 + \tau_{32}\rho_2 + \Delta\rho_3$ for state 3 since the terms associated with the vector $\boldsymbol{e}$ are not state dependent.

We can examine inequalities between the variances as follows:

$$v_1 > v_2 \Leftrightarrow \Delta\rho_1 + \tau_{12}\rho_2 + \tau_{13}\rho_3 > \tau_{21}\rho_1 + \Delta\rho_2 + \tau_{23}\rho_3$$
$$\Leftrightarrow (\Delta - \tau_{21})\rho_1 + (\tau_{12} - \Delta)\rho_2 + (b_3 - a_3)\rho_3 > 0.$$

Similarly $v_2 > v_3 \Leftrightarrow (c_1 - b_1)\rho_1 + (\Delta - \tau_{32})\rho_2 + (\tau_{23} - \Delta)\rho_3 > 0$,

and $v_3 > v_1 \Leftrightarrow (\tau_{31} - \Delta)\rho_1 + (a_2 - c_2)\rho_2 + (\Delta - \tau_{13})\rho_3 > 0$.

To choose a starting state to minimise the variance of the mixing time, select state 1, for example, if both $v_1 \leq v_2$ and $v_1 \leq v_3$.

Because of the high dimension of the parameter space (a_2, a_3, b_1, b_3, c_1 and c_2), the verification and establishment of regions where such general inequalities are satisfied are better considered in special cases. We look at a selection of the special cases considered in the earlier paper [3].

Case 1*: "Minimal period 3"*

If $a_2 = b_3 = c_1 = 1$, the Markov chain is periodic, period 3, with transitions occurring 1 $\rightarrow 2 \rightarrow 3 \rightarrow 1$ ... and transition matrix $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

In this particular case, the distribution of the mixing time has a simple derivation. In particular, given $X_0 = i$, $T = k$, (k = 1, 2, 3) with probability 1/3, for all such i. (Each of the states 1, 2, or 3 are sampled according to the stationary distribution of the Markov chain – each with probability 1/3. Since the chain cycles through all the states, the state sampled will be reached from the starting state i in either 1, 2 or 3 steps (each with the same probability of 1/3*).* Thu*s*

$$E(T \mid X_0 = i) = 1.(1/3) + 2.(1/3) + 3.(1/3) = 2.$$

Further

$$\mathrm{var}(T \mid X_0 = i) = (1-2)^2(1/3) + (2-2)^2(1/3) + (3-2)^2(1/3) = 2/3 = 0.6667\,.$$

This is consistent with the general theory developed above and verified below.
Note that$\Delta_1 = \Delta_2 = \Delta_3 = 1$, $\Delta = 3$, implying $\boldsymbol{\pi}^T$ = (1/3, 1/3. 1/3).
Further $\tau_{12} = 1$, $\tau_{13} = 2$, $\tau_{21} = 2$, $\tau_{23} = 1$, $\tau_{31} = 1$, $\tau_{32} = 2$, $\tau = 3$ with

$$M = \begin{bmatrix} 3 & 1 & 2 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix},\ L = \begin{bmatrix} 1 & 1/3 & 2/3 \\ 2/3 & 1 & 1/3 \\ 1/3 & 2/3 & 1 \end{bmatrix} \text{ leading to } \eta = 2.$$

Further taking $G = [I - P + \boldsymbol{e}\boldsymbol{e}_1^T]^{-1}$, $\kappa_1 = 3$, $\kappa_2 = 3$, $\kappa_3 = 3$, $\rho_1 = 3$, $\rho_2 = 3$, $\rho_3 = 3$ with

$$M^{(2)} = \begin{bmatrix} 9 & 1 & 4 \\ 4 & 9 & 1 \\ 1 & 4 & 9 \end{bmatrix}, L^{(2)} == \begin{bmatrix} 3 & 1/3 & 4/3 \\ 4/3 & 3 & 1/3 \\ 1/3 & 4/3 & 3 \end{bmatrix} \quad \text{leading to}$$

$$\boldsymbol{\eta}^{(2)} = \begin{bmatrix} 4\ 2/3 \\ 4\ 2/3 \\ 4\ 2/3 \end{bmatrix},\ \boldsymbol{\nu} = \begin{bmatrix} 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}.$$

Note also that $g = 1$, $tr(G) = 2$, $tr(G^2) = 4/3$, $\boldsymbol{\alpha}^T = (2, 2, 2)$ with Eqns.(4.5) and (4.6) yielding the above results for $\boldsymbol{\eta}^{(2)}$ and $\boldsymbol{\nu}$.
In this case, the symmetry of the transition matrix does not provide any opportunity for $\boldsymbol{\eta}^{(2)}$ or $\mathbf{v}$ to vary according to the starting state $X_0 = i$, (i = 1, 2 or 3). Further, the values of η and $\mathbf{v}$ are the smallest possible amongst all 3-state Markov chains.

Case 2: "Period 2"
If $a_2 = 1$, $b_1 + b_3 = 1$, $c_2 = 1$, the Markov chain is periodic, period *2* (with transitions alternating between the states {1, 3} and {2}), and transition matrix $P = \begin{bmatrix} 0 & 1 & 0 \\ b_1 & 0 & b_3 \\ 0 & 1 & 0 \end{bmatrix}$.

Then $\Delta_1 = b_1$, $\Delta_2 = 1$, $\Delta_3 = b_3$, $\Delta = 2$, implying $\boldsymbol{\pi}^T = (b_1/2, 1/2, b_3/2)$.
Further $\tau_{12} = 1$, $\tau_{13} = 2$, $\tau_{21} = b_3 + 1$, $\tau_{23} = b_1 + 1$, $\tau_{31} = 2$, $\tau_{32} = 1$, $\tau = 3$ with

$$M=\begin{bmatrix} 2/b_1 & 1 & 2/b_3 \\ (1+b_3)/b_1 & 2 & (1+b_1)/b_3 \\ 2/b_1 & 1 & 2/b_3 \end{bmatrix},\ L=\begin{bmatrix} 1 & 1/2 & 1 \\ (1+b_3)/2 & 1 & (1+b_1)/2 \\ 1 & 1/2 & 1 \end{bmatrix}$$

leading to $\eta = 2.5$.

Further taking $G=[I-P+\boldsymbol{e}\boldsymbol{e}_1^T]^{-1}$, $\kappa_1 = 1 + 3b_3$, $\kappa_2 = 1$, $\kappa_3 = 1 + 3b_1$, $\rho_1 = (1 + 3b_3)/b_1$, $\rho_2 = 1$, $\rho_3 = (1 + 3b_1)/b_3$, with

$$M^{(2)}=\begin{bmatrix} 4(1+b_3)/b_1^2 & 1 & 4(1+b_1)/b_3^2 \\ (1+6b_3+b_3^2)/b_1^2 & 4 & (1+6b_1+b_1^2)/b_3^2 \\ 4(1+b_3)/b_1^2 & 1 & 4(1+b_1)/b_3^2 \end{bmatrix},$$

$$L^{(2)}=\begin{bmatrix} 2(1+b_3)/b_1 & 1/2 & 2(1+b_1)/b_3 \\ (1+6b_3+b_3^2)/2b_1 & 2 & (1+6b_1+b_1^2)/2b_3 \\ 2(1+b_3)/b_1 & 1/2 & 2(1+b_1)/b_3 \end{bmatrix},$$

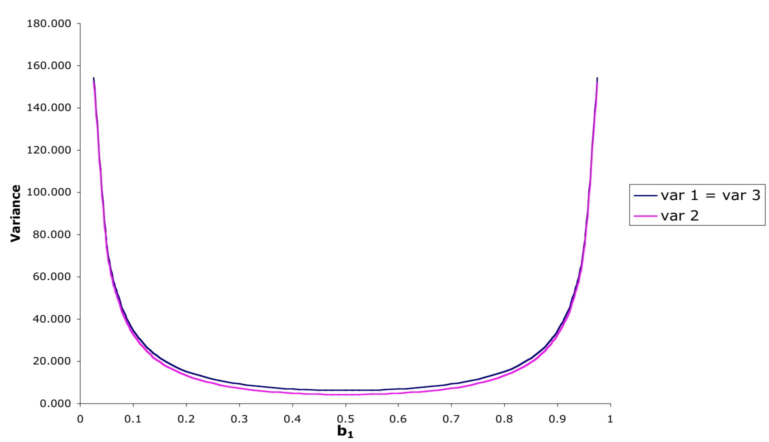
leading, after simplification using $b_1+b_3=1, b_1^2+b_3^2=1-2b_1b_3, b_1^3+b_3^3=1-3b_1b_3$, to

$$\boldsymbol{\eta}^{(2)}=\frac{4}{b_1b_3}\begin{bmatrix}1\\1\\1\end{bmatrix}-\begin{bmatrix}3.5\\5.5\\3.5\end{bmatrix} \text{and } \boldsymbol{v}=\begin{bmatrix}var_1\\var_2\\var_3\end{bmatrix}=\frac{4}{b_1b_3}\begin{bmatrix}1\\1\\1\end{bmatrix}-\begin{bmatrix}9.75\\11.75\\9.75\end{bmatrix}.$$

Note also that $g = 1$, $tr(G) = 2.5$, $tr(G^2) = 2.25$, $\boldsymbol{\alpha}^T=((3+b_3)/2b_1, 3/2, (3+b_1)/2b_3)$ with Eqns.(4.5) and (4.6) yielding the above results for $\boldsymbol{\eta}^{(2)}$ and $\boldsymbol{v}$.

Graph 5 gives a plot of the variances of the mixing times starting in states 1, 2 and 3.

One notes that for the same b_1(and b_3), $var_2 = var_1 + 2 = var_3 + 2$, thus if one wishes to minimise the variance of the mixing time one should always start in state 2.



Graph 5: Variance of Mixing Times for Case 2

Case 3: "Constant movement"
If $a_2 + a_3 = 1$, $b_1 + b_3 = 1$, $c_1 + c_2 = 1$, then at each step the chain does not remain at the state but moves to one of the other states. The Markov chain is irreducible. It is regular if $0 < a = a_2 < 1$, $0 < b = b_3 < 1$, $0 < c = c_1 < 1$. The transition matrix is

$$P = \begin{bmatrix} 0 & a & 1-a \\ 1-b & 0 & b \\ c & 1-c & 0 \end{bmatrix}.$$

Now $\Delta_1 = 1 - b(1 - c)$, $\Delta_2 = 1 - c(1 - a)$, $\Delta_3 = 1 - a(1 - b)$, $\tau_{12} = 2 - a$, $\tau_{13} = 1 + a$, $\tau_{21} = 1 + b$, $\tau_{23} = 2 - b$, $\tau_{31} = 2 - c$, $\tau_{32} = 1 + c$, implying that $\Delta = 3 - b(1 - c) - c(1 - a) - a(1 - b)$, $\tau = 3$, and hence that $\eta = 1 + 3/[3 - b(1-c) - c(1-a) - a(1-b)]$.

In [3] we showed $2 \le \eta \le 2.5$. The minimal value of 2 occurs when $a = b = c = 1$, and this case reduces to the "period 3" Case 1 above. Further when $a = b = c = 0$ this case again reduces to a periodic, "period 3" chain but with transitions $1 \to 3 \to 2 \to 1$ .... .

The maximal value of 2.5 occurs when any pair of (a, b, c) take the values 0 and 1, say $a = 1$, $c = 0$ when this case reduces to the "period *2*" Case *2* above. For the regular case $2 < \eta < 2.5$.

Expression (4.5) gives explicit expressions for $\mathbf{v}$, involving terms $\boldsymbol{\beta}$, with $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \beta_3)$. Since $\beta_i - \beta_j = \frac{\Delta^2}{2}(v_i - v_j)$, we can deduce inequalities between the variances of the mixing times starting in state i, v_I, without finding expressions for the constant terms.

It is difficult to give simple forms for the β_i, and the differences between the β_i. We have however computed values of β_i for a, b and $c = 0.1(0.1)0.9$, and determined the particular state i where the minimum β_i, and hence v_i, occurs. For each such *a, b, c* combination, Table 1 displays the relevant starting state. The entry 123 indicates that each starting state gives the same variance.

Note that the regions for the appropriate starting state are not simple in form. One can make some simple observations from Table 1.

If $a \ge b + 0.2$ then either $v_1 < v_3$ or $v_2 < v_3$ so that the minimum variance cannot occur at state 3 so that state 3 would not be a recommended starting state for the mixing process. Similarly if $b \ge c + 0.2$ do not start the mixing in state 1 while if if $c \ge a + 0.2$ do not start in state 2.

In any given Case 3 situation, by relabelling the states if necessary, effectively only four different scenarios arise *(i)* $a > b > c$ *(ii)* $a = b > c$ *(iii)* $a > b = c$ *(iv)* $a = b = c$. For *(i)* and *(ii)* never start in state 1; for *(iii)* always start in state 2; while for *(iv)* the start state is immaterial. We have not been able to derive simple rules based on the values of a, b and c to establish the recommended starting state other than using the guidelines evident in Table 1.

c = 0.1

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	123	3	3	3	3	3	3	3	3
0.2	2	2	2	2	3	3	3	3	3
0.3	2	2	2	2	2	2	3	3	3
0.4	2	2	2	2	2	2	3	3	3
0.5	2	2	2	2	2	2	2	3	3
0.6	2	2	2	2	2	2	2	3	3
0.7	2	2	2	2	2	2	2	3	3
0.8	2	2	2	2	2	2	2	3	3
0.9	2	2	2	2	2	2	2	3	3

c = 0.2

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	3	3	3	3	3	3	3	3
0.2	1	123	3	3	3	3	3	3	3
0.3	1	2	2	2	3	3	3	3	3
0.4	1	2	2	2	2	3	3	3	3
0.5	2	2	2	2	2	2	3	3	3
0.6	2	2	2	2	2	2	3	3	3
0.7	2	2	2	2	2	2	2	3	3
0.8	2	2	2	2	2	2	2	3	3
0.9	2	2	2	2	2	2	2	3	3

c = 0.3

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	3	3	3	3	3	3	3	3
0.2	1	1	3	3	3	3	3	3	3
0.3	1	1	123	3	3	3	3	3	3
0.4	1	1	2	2	3	3	3	3	3
0.5	1	2	2	2	2	3	3	3	3
0.6	1	2	2	2	2	2	3	3	3
0.7	2	2	2	2	2	2	3	3	3
0.8	2	2	2	2	2	2	2	3	3
0.9	2	2	2	2	2	2	2	3	3

c = 0.4

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	3	3	3	3	3	3	3	3
0.2	1	1	3	3	3	3	3	3	3
0.3	1	1	1	3	3	3	3	3	3
0.4	1	1	1	123	3	3	3	3	3
0.5	1	1	2	2	2	3	3	3	3
0.6	1	2	2	2	2	2	3	3	3
0.7	2	2	2	2	2	2	3	3	3
0.8	2	2	2	2	2	2	2	3	3
0.9	2	2	2	2	2	2	2	3	3

c = 0.5

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	1	3	3	3	3	3	3	3
0.2	1	1	1	3	3	3	3	3	3
0.3	1	1	1	1	3	3	3	3	3
0.4	1	1	1	1	3	3	3	3	3
0.5	1	1	1	1	123	3	3	3	3
0.6	1	1	2	2	2	2	3	3	3
0.7	1	2	2	2	2	2	3	3	3
0.8	2	2	2	2	2	2	2	3	3
0.9	2	2	2	2	2	2	2	3	3

c = 0.6

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	1	3	3	3	3	3	3	3
0.2	1	1	1	1	3	3	3	3	3
0.3	1	1	1	1	1	3	3	3	3
0.4	1	1	1	1	1	3	3	3	3
0.5	1	1	1	1	1	3	3	3	3
0.6	1	1	1	1	1	123	3	3	3
0.7	1	2	2	2	2	2	3	3	3
0.8	2	2	2	2	2	2	2	3	3
0.9	2	2	2	2	2	2	2	3	3

c = 0.7

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	1	1	1	3	3	3	3	3
0.2	1	1	1	1	1	1	3	3	3
0.3	1	1	1	1	1	1	1	3	3
0.4	1	1	1	1	1	1	1	3	3
0.5	1	1	1	1	1	1	1	3	3
0.6	1	1	1	1	1	1	1	3	3
0.7	1	1	2	2	2	2	123	3	3
0.8	2	2	2	2	2	2	2	3	3
0.9	2	2	2	2	2	2	2	2	3

c = 0.8

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	1	1	1	1	1	1	1	1
0.2	1	1	1	1	1	1	1	1	1
0.3	1	1	1	1	1	1	1	1	1
0.4	1	1	1	1	1	1	1	1	1
0.5	1	1	1	1	1	1	1	1	1
0.6	1	1	1	1	1	1	1	1	1
0.7	1	1	1	1	1	1	1	1	3
0.8	2	2	2	2	2	2	2	123	3
0.9	2	2	2	2	2	2	2	2	3

c = 0.9

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	1	1	1	1	1	1	1	1	1
0.2	1	1	1	1	1	1	1	1	1
0.3	1	1	1	1	1	1	1	1	1
0.4	1	1	1	1	1	1	1	1	1
0.5	1	1	1	1	1	1	1	1	1
0.6	1	1	1	1	1	1	1	1	1
0.7	1	1	1	1	1	1	1	1	1
0.8	2	2	2	2	2	2	1	1	1
0.9	2	2	2	2	2	2	2	2	123

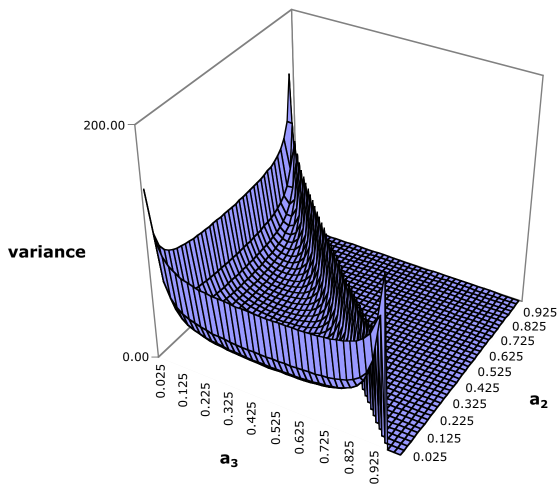
Table 1: States where minimum variance of mixing time occurs for Case 3.

Case 4: "Independent"

Let $b_1 = c_1 = 1 - a_2 - a_3$, $c_2 = a_2 = 1 - b_1 - b_3$, $b_3 = a_3 = 1 - c_1 - c_2$, implying that the Markov chain is equivalent to independent trials on the state space $S = \{1, 2, 3\}$.

$\Delta_1 = 1 - a_2 - a_3$, $\Delta_2 = a_2$, $\Delta_3 = a_3$, $\Delta = 1$. For all i, j, $\tau_{ij} = 1$, $\tau = 1$, $\kappa_1 = a_2 + a_3$, $\kappa_2 = 1 - a_2$, $\kappa_3 = 1 - a_3$, with $\rho_i = \kappa_i/\Delta_i$ implying $\rho_1 = \dfrac{a_2 + a_3}{1 - a_2 - a_3}$, $\rho_2 = \dfrac{1 - a_2}{a_2}$, and $\rho_2 = \dfrac{1 - a_3}{a_3}$.

It is easily shown that $\eta = 3$, $\boldsymbol{\eta}^{(2)} = \{3 + 2(\rho_1 + \rho_2 + \rho_3)\}\boldsymbol{e}$, and $\boldsymbol{\nu} = 2\{\rho_1 + \rho_2 + \rho_3 - 3\}\boldsymbol{e}$. As to be expected, the variances of the mixing times starting in any state are constant with $\nu_i = 2\{\rho_1 + \rho_2 + \rho_3 - 3\} \geq 6$. The minimum variance of 6 occurs when $a_2 = a_3 = 1/3$.



Graph 6: Variance of Mixing Times for Case 4

Case 5: "Cyclic drift"
Let, $a_3 = b_1 = c_2 = 0$, $a_2 = a$, $b_3 = b$, $c_1 = c$ with $0 < a < 1$, $0 < b < 1$, $0 < c < 1$, implying that the Markov chain is regular with transition matrix $P = \begin{bmatrix} 1-a & a & 0 \\ 0 & 1-b & b \\ c & 0 & 1-c \end{bmatrix}$. At each transition the chain either remains in the same state i or moves to state $i + 1$ (or 1 if $i = m$).
Now $\Delta_1 = bc$, $\Delta_2 = ca$, $\Delta_3 = ab$, $\Delta = bc + ca + ab$, 1, $\tau_{12} = c$, $\tau_{13} = a + b$, $\tau_{21} = b + c$, $\tau_{23} = a$, $\tau_{31} = b$, $\tau_{32} = a + c$, leading to $\tau = \mathrm{a + b + c}$. Further $\kappa_1 = a\{bc + b^2 + c^2\}$, $\kappa_2 = b\{ac + a^2 + c^2\}$, $\kappa_3 = c\{ab + a^2 + b^2\}$. Note that $0 < \tau < 3$ and $0 < \Delta < 3$ implying $\eta = 1 + \dfrac{\tau}{\Delta} = 1 + \dfrac{a+b+c}{bc+ca+ab}$.
When $a + b + c \to 3$ then $bc + ca + ab \to 3$ and $\eta \to 2$ (as in Case 1).
When $a + b + c \to 0$ then $bc + ca + ab \to 0$, but the behaviour of η depends upon the rates of convergence and can be large. In this situation the Markov chain resides for a large number of transitions in each state so that there is little movement implying that the mixing time can become excessively large.

For any situation the mean time to mixing does not depend on the starting state, but we explore scenarios in order to determine the appropriate starting state under the condition that the variance of the mixing time is minimised. We use expression (4.5) giving explicit expressions for $\mathbf{v}$ in terms of elements of $\boldsymbol{\beta}$.

As in Case 3, it is not easy to give explicit simple conditions on the parameters *a*, *b*, *c* to describe the region where inequalities between the β_i occur. It can be shown that $\rho_1 < \rho_2 \Leftrightarrow a < b,\ \rho_2 < \rho_3 \Leftrightarrow b < c,\ \rho_3 < \rho_1 \Leftrightarrow c > a.$ Further

$$\beta_1 - \beta_2 = \frac{\Delta}{abc}\left\{\Delta c(a-b) + c(b^2 - ac) + b(c^2 - ab)\right\} = \frac{\Delta^2}{2}(\nu_1 - \nu_2).$$

Conditions under which $\beta_1 - \beta_2 > 0$ are not easy to express. Values of β_i for *a*, *b* and *c* = 0.1(0.1)0.9, can be easily computed and the particular state *i* where the minimum β_i, and hence ν_i, occurs, determined. For each such *a, b, c* combination, Table 2 displays the relevant starting state. Where multiple entries are given any such listed state can be used as they provide the same minimum variance.

Note that the regions for the appropriate starting state are not simple in form but some inequalities can be deduced, from the calculations underpinning Table 2, to assist in the derivation of some "rules of thumb" for determining the starting state.

If $a < b$ then either $\nu_1 < \nu_3$ or $\nu_2 < \nu_3$ which implies that one should not start in state 3. Similarly, if $b < c$, never start in state 1 while if $c < a$, never start in state 2.

If $\Delta - a \leq -0.09$ then either $\nu_1 < \nu_2$ or $\nu_3 < \nu_2$ which implies that one should not start in state 2. Similarly, if $\Delta - b \leq -0.09$, never start in state 3 while if $\Delta - a \leq -0.09$, don't start in state 1.

If $a^2 - bc \leq -0.10$ then either $\nu_1 < \nu_3$ or $\nu_2 < \nu_3$ implying that one should not start in state 3. Further, if $b^2 - ac \leq -0.10$, never start in state 1 while if $c^2 - ab \leq -0.10$, don't start in state 2.

Note also that $b^2 - ac > 0$ and $c^2 - bc > 0 \Rightarrow a - b > 0.$

As for Case 3, in this situation, by relabelling the states if necessary, effectively only four different scenarios arise *(i)* $a > b > c$ *(ii)* $a = b > c$ *(iii)* $a > b = c$ *(iv)* $a = b = c$. For *(i)* never start in state 3; for *(ii)* never start in state 2; for *(iii)* never start in state 2 ("Rule of thumb": start in state 1 if Δ < b or start in state 3 if Δ > b, either 1 or 3 if Δ = b); while for *(iv)* the start state is immaterial. There are no simple universal rules to determine the starting state in *(ii)* and *(iii).*

Note that changes in the parameter values in different directions can lead to quite dramatic changes especially around the parameter set (a, b, c) = (0.5, 0.5, 0.5) where it is immaterial which state one starts in. Increase, or decrease any of the parameters individually only slightly can lead to quite different recommended starting points. For (0.51, 0.5, 0.5) start in state 3, for (0.5, 0.51, 0.5) start in state 1, for (0.5, 0.5,0.51) start in state 2, while for (0.49, 0.5, 0.5) start in state 2, for (0.5, 0.49, 0.5) start in state 3, and for (0.5, 0.5, 0.49) start in state 1. Thus there is a great deal sensitivity around the central parameter set (0.5, 0.5, 0.5). In general, this dramatic shift in recommended starting states is also exhibited around each of the equal valued parameter sets (*a*, *a*, *a*).

One typically assumes some sense of stability in the Markov chain around such a parameter set where the stationary distribution displays stability with $\boldsymbol{\pi}^T = (1/3,\ 1/3,\ 1/3)$ and mean time to mixing of $\eta = 1 + (1/a)$.

c = 0.1

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	123	2	2	2	1	1	1	1	2
0.2	1	1	1	1	1	1	1	1	1
0.3	1	1	1	1	1	1	1	1	1
0.4	1	1	1	1	1	1	1	1	1
0.5	3	1	1	1	1	1	1	1	1
0.6	3	1	1	1	1	1	1	1	1
0.7	3	1	1	1	1	1	1	1	1
0.8	3	1	1	1	1	1	1	1	1
0.9	3	1	1	1	1	1	1	1	1

c = 0.2

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	3	2	2	2	2	2	2	2	2
0.2	3	123	2	12	1	1	1	1	1
0.3	3	1	1	1	1	1	1	1	1
0.4	3	13	1	1	1	1	1	1	1
0.5	3	3	1	1	1	1	1	1	1
0.6	3	3	1	1	1	1	1	1	1
0.7	3	3	1	1	1	1	1	1	1
0.8	3	3	1	1	1	1	1	1	1
0.9	3	3	1	1	1	1	1	1	1

c = 0.3

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	3	2	2	2	2	2	2	2	2
0.2	3	3	2	2	2	2	2	2	2
0.3	3	3	123	1	1	1	1	1	1
0.4	3	3	3	1	1	1	1	1	1
0.5	3	3	3	1	1	1	1	1	1
0.6	3	3	3	1	1	1	1	1	1
0.7	3	3	3	1	1	1	1	1	1
0.8	3	3	3	1	1	1	1	1	1
0.9	3	3	3	3	1	1	1	1	1

c = 0.4

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	3	2	2	2	2	2	2	2	2
0.2	3	23	2	2	2	2	2	2	2
0.3	3	3	2	2	2	2	2	2	1
0.4	3	3	3	123	1	1	1	1	1
0.5	3	3	3	3	1	1	1	1	1
0.6	3	3	3	3	1	1	1	1	1
0.7	3	3	3	3	3	1	1	1	1
0.8	3	3	3	3	3	1	1	1	1
0.9	3	3	3	3	3	13	1	1	1

c = 0.5

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2	2	2	2	2	2	2	2	2
0.2	3	2	2	2	2	2	2	2	2
0.3	3	3	2	2	2	2	2	2	2
0.4	3	3	3	2	2	2	1	1	1
0.5	3	3	3	3	123	1	1	1	1
0.6	3	3	3	3	3	1	1	1	1
0.7	3	3	3	3	3	3	1	1	1
0.8	3	3	3	3	3	3	3	1	1
0.9	3	3	3	3	3	3	3	3	1

c = 0.6

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2	2	2	2	2	2	2	2	2
0.2	3	2	2	2	2	2	2	2	2
0.3	3	3	2	2	2	2	2	2	2
0.4	3	3	3	2	2	2	2	2	12
0.5	3	3	3	3	2	2	1	1	1
0.6	3	3	3	3	3	123	1	1	1
0.7	3	3	3	3	3	3	1	1	1
0.8	3	3	3	3	3	3	3	13	1
0.9	3	3	3	3	3	3	3	3	3

c = 0.7

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2	2	2	2	2	2	2	2	2
0.2	3	2	2	2	2	2	2	2	2
0.3	3	3	2	2	2	2	2	2	2
0.4	3	3	3	2	2	2	2	2	2
0.5	3	3	3	2	2	2	2	1	1
0.6	3	3	3	3	2	2	2	1	1
0.7	3	3	3	3	3	3	123	1	1
0.8	3	3	3	3	3	3	3	3	1
0.9	3	3	3	3	3	3	3	3	3

c = 0.8

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2	2	2	2	2	2	2	2	2
0.2	3	2	2	2	2	2	2	2	2
0.3	3	3	2	2	2	2	2	2	2
0.4	3	3	3	2	2	2	2	2	2
0.5	3	3	3	2	2	2	2	2	1
0.6	3	3	3	3	2	2	2	12	1
0.7	3	3	3	3	2	2	2	1	1
0.8	3	3	3	3	3	23	2	123	1
0.9	3	3	3	3	3	3	3	3	1

c = 0.9

a \ b	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	2	2	2	2	2	2	2	2	2
0.2	3	2	2	2	2	2	2	2	2
0.3	3	3	2	2	2	2	2	2	2
0.4	3	3	2	2	2	2	2	2	2
0.5	3	3	3	2	2	2	2	2	2
0.6	3	3	3	23	2	2	2	2	1
0.7	3	3	3	3	2	2	2	2	1
0.8	3	3	3	3	2	2	2	2	1
0.9	3	3	3	3	3	2	2	2	123

Table 2: States where minimum variance of mixing time occurs for Case 5.

References

[1] Heyman D.P. and D. P. O'Leary D.P. (1995). What is Fundamental for Markov chains? First passage times, Fundamental matrices, and group generalized Inverses, in *Computations with Markov chains, Proceedings of the 2nd International Workshop on the Numerical Solution of Markov Chains*, (Stewart W.J. Ed), Kluwer Academic Publishers, Dordrecht, 151-159.

[2] Heyman D. P. and A. Reeves A (1989). Numerical solution of linear equations arising in Markov chain models. *ORSA J. Comput*. 1: 52 – 60.

[3] Hunter J. J. (2006). Mixing times with applications to perturbed Markov chains, *Linear Algebra Appl*. (Available on-line)

[4] Hunter J. J. (2005a). Simple procedures for finding mean first passage times in Markov chains, *Proceedings of IWMS 2005, Res. Lett. Inf. Math. Sci.,* 8: 209 – 226.

[5] Hunter J. J. (2005b) Stationary distributions and mean first passage times of perturbed Markov chains, *Linear Algebra Appl.* 410: 217-243.

[6] Hunter J. J. (1992). Stationary distributions and mean first passage times in Markov chains using generalized inverses, *Asia-Pacific J. Oper. Res.* 9: 145-153.

[7] Hunter J. J. (1990). Parametric forms for generalized inverses of Markovian kernels and their applications, *Linear Algebra Appl*. 127: 71-84.

[8] Hunter J. J. (1988). Characterizations of generalized inverses associated with Markovian kernels, *Linear Algebra Appl.* 102: 121-142.

[9] Hunter J. J. (1983). *Mathematical Techniques of Applied Probability, Volume 1, Discrete Time Models: Basic Theory*, Academic, New York.

[10] Hunter J. J. (1983). *Mathematical Techniques of Applied Probability, Volume 2, Discrete Time Models: Techniques and Applications*, Academic, New York.

[11] Hunter J. J. (1982): Generalized inverses and their application to applied probability problems, *Linear Algebra Appl.* 45: 157-198.

[12] Hunter J. J. (1969). On the moments of Markov renewal processes, *Adv. Appl. Probab*. 1: 188-210.

[13] Kemeny J. G. and Snell J. L. (1960). *Finite Markov Chains*, Van Nostrand, New York.

[14] Meyer C. D. Jr. (1975). The role of the group generalized inverse in the theory of finite Markov chains, *SIAM Rev*. 17: 443-464.